

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 07:25:04 ; Search time 31.2046 Seconds
(without alignments)
2218.735 Million cell updates/sec

Title: US-10-030-937-8

Perfect score: 193

Sequence: 1 MQSLMQAPLLIALGLLILATP.....LSSGKRLGCIKIASLKI 193

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Genesep 23Sep04: *

1: genesep1980s: *

2: genesep1990s: *

3: genesep2000s: *

4: genesep2001s: *

5: genesep2002s: *

6: genesep2003as: *

7: genesep2003bs: *

8: genesep2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	193	100.0	193	4	AAB31901	Aab31901 Amino aci
2	193	100.0	193	4	AAB31902	Aab31902 Amino aci
3	193	100.0	193	4	AAB31904	Aab31904 Amino aci
4	193	100.0	193	4	AAB31896	Aab31896 Amino aci
5	193	100.0	193	4	AAB31928	Aab31928 Amino aci
6	193	100.0	193	4	AAB31903	Aab31903 Amino aci
7	193	100.0	193	5	ABG311345	Abg311345 Human GM2
8	193	100.0	193	5	ABP65212	Abp65212 Hypoxia-r
9	193	100.0	193	7	ADN95858	Adn95858 Human BPC
10	193	100.0	193	8	ADN03620	Adn03620 Antipsori
11	189	97.9	189	4	AAB31900	Aab31900 Amino aci
12	178	92.2	178	4	AAB31898	Aab31898 Amino aci
13	174	90.2	174	4	ABG00720	Abg00720 Novel hum
14	174	90.2	174	8	ADQ17712	Adq17712 Human sof
15	162	83.9	178	5	ABG31346	Abg31346 Non-glyco
16	124	64.2	200	4	AAB31899	Aab31899 Amino aci
17	112	58.0	193	4	AAB31897	Aab31897 Amino aci
18	75	38.9	76	4	ABG00717	Abg00717 Novel hum
19	61	31.6	61	4	AM15082	Aam15082 Peptide #
20	61	31.6	61	4	ABE34073	Abbe34073 Peptide #
21	61	31.6	61	4	AM27530	Aam27530 Peptide #
22	61	31.6	61	4	ABE28897	Abbe28897 Peptide #
23	61	31.6	61	4	ABE19511	Abbe19511 Protein #
24	61	31.6	61	4	AM67236	Aam67236 Human bon
25	61	31.6	61	4	AM54855	Aam54855 Human bra

26	61	31.6	61	4	ABG48902	Abg48902 Human liv
27	61	31.6	61	4	AM02818	Aam02818 Peptide #
28	61	31.6	61	5	ABG36887	Abg36887 Human pep
29	18	9.3	18	4	AAB31920	Aab31920 Amino aci
30	18	9.3	199	2	AAW10656	Aaw10656 Rat GM2 a
31	16	8.3	131	4	ABG00719	Abg00719 Novel hum
32	15	7.8	15	4	AAB31918	Aab31918 Amino aci
33	13	6.7	13	4	AAB31919	Aab31919 Amino aci
34	11	5.7	11	4	AAU25350	Aau25350 Schizophr
35	11	5.7	11	4	AAU15694	Aau15694 Schizophr
36	11	5.7	11	8	ADO78961	Ado78961 Schizophr
37	9	4.7	16	4	AAB31926	Aab31926 Amino aci
38	9	4.7	262	2	AAW55111	Aaw55111 Streptoco
39	9	4.7	262	2	ABP54605	Abp54605 S. pneumo
40	9	4.7	262	7	ADC45179	Adc45179 S. pneumo
41	9	4.7	270	2	AAW61241	Aaw61241 Streptoco
42	9	4.7	270	5	ABP54659	Abp54659 S. pneumo
43	9	4.7	270	7	ADC45287	Adc45287 S. pneumo
44	9	4.7	291	6	ABU02596	Abu02596 S. pneumo
45	9	4.7	291	8	ADK47686	Adk47686 Streptoco

ALIGNMENTS

RESULT 1
AAB31901
ID AAB31901 standard; protein; 193 AA.

XX AAB31901;

DT 15-MAY-2001 (first entry)

DE Amino acid sequence of a human protein.

KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

OS Homo sapiens.

PN WO200105422-A2.

PD 25-JAN-2001.

PF 17-JUL-2000; 2000WO-FR002057.

PR 15-JUL-1999; 99FR-00009372.

XX (INNR) BIOMERIEUX STELHYS.

PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

DR WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.

PS Claim 1; Page 162-163; 209pp; French.

XX The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and

CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;

Query Match 100.0%; Score 193; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 9e-187;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSILMQAPLIIALGILLATPAQHLKPSQSSFWDCDEGKDPVIRSLTLEPDPV 60
DB 1 MQSILMQAPLIIALGILLATPAQHLKPSQSSFWDCDEGKDPVIRSLTLEPDPV 60
QY 61 PGNVTLVVGSTVSPSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
DB 61 PGNVTLVVGSTVSPSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
QY 121 TGBPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180
DB 121 TGBPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180
QY 181 LGCTIKIAASLKI 193
DB 181 LGCTIKIAASLKI 193

RESULT 2
AAB31902
ID AAB31902 standard; protein; 193 AA.

XX AC AAB31902;
XX AC
XX DT 15-MAY-2001 (first entry)
XX DE Amino acid sequence of a human protein.
XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
XX KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
XX KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
XX KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
XX KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX OS Homo sapiens.
XX PN WO200105422-A2.
XX PD 25-JAN-2001.
XX PF 17-JUL-2000; 2000WO-FR002057.
XX PR 15-JUL-1999; 99FR-00009372.
XX PA (INNR) BIOMERIEUX STELHYS.
XX PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX DR WPI; 2001-159475/16.
XX PT Detecting, preventing and treating degenerative, neurological and
XX PT autoimmune diseases, particularly multiple sclerosis, using specified
XX PT polypeptides or related nucleic acid or ligand.
XX PS Claim 1; Page 163; 209pp; French.

XX CC The present sequence represents a human protein, which is used in the
XX CC method of the invention. The specification describes a method which uses
XX CC at least one polypeptide or polynucleotide sequence belonging to the
XX CC perlecan, precursor of the retinol-binding plasma protein, precursor of
XX CC the ganglioside GM2 activator, calgranulin B or saposin B protein
XX CC families. The method is used for detecting, preventing or treating a

CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;

Query Match 100.0%; Score 193; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 9e-187;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQSILMQAPLIIALGILLATPAQHLKPSQSSFWDCDEGKDPVIRSLTLEPDPV 60
DB 1 MQSILMQAPLIIALGILLATPAQHLKPSQSSFWDCDEGKDPVIRSLTLEPDPV 60
QY 61 PGNVTLVVGSTVSPSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
DB 61 PGNVTLVVGSTVSPSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
QY 121 TGBPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180
DB 121 TGBPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180
QY 181 LGCTIKIAASLKI 193
DB 181 LGCTIKIAASLKI 193

RESULT 3
AAB31904
ID AAB31904 standard; protein; 193 AA.

XX AC AAB31904;
XX DT 15-MAY-2001 (first entry)
XX DE Amino acid sequence of a human protein.
XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
XX KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
XX KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
XX KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
XX KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX OS Homo sapiens.
XX PN WO200105422-A2.
XX PD 25-JAN-2001.
XX PF 17-JUL-2000; 2000WO-FR002057.
XX PR 15-JUL-1999; 99FR-00009372.
XX PA (INNR) BIOMERIEUX STELHYS.
XX PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX DR WPI; 2001-159475/16.
XX PT Detecting, preventing and treating degenerative, neurological and
XX PT autoimmune diseases, particularly multiple sclerosis, using specified
XX PT polypeptides or related nucleic acid or ligand.
XX PS Claim 1; Page 164-165; 209pp; French.
XX CC The present sequence represents a human protein, which is used in the
XX CC method of the invention. The specification describes a method which uses

CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 100.0%; Score 193; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 9e-187;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MOSLMAQPLLIAGLLLATPAQAHLLKPSQSFSDNCEGKDPVIRSLTLEDDPIV 60
 DB 1 MOSLMAQPLLIAGLLLATPAQAHLLKPSQSFSDNCEGKDPVIRSLTLEDDPIV 60
 QY 61 PGNVTLVVGVSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYGCTFEHFCVDLMDLIP 120
 DB 61 PGNVTLVVGVSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYGCTFEHFCVDLMDLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLEPLSWLTTCNYRIESVLSGGR 180
 DB 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLEPLSWLTTCNYRIESVLSGGR 180
 QY 181 LGCIKIAASLKI 193
 DB 181 LGCIKIAASLKI 193

RESULT 4
 AAB31896
 ID AAB31896 standard; protein; 193 AA.
 AC AAB31896;
 DT 15-MAY-2001 (first entry)
 DE Amino acid sequence of the human ganglioside GM2 activator protein.
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 FN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INNR) BIOMERIEUX STELHYS.
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI: 2001-159475/16.
 DR N-PSDB; AAF54698.
 XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX
 PS Claim 23; Page 158-159; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX

SQ Sequence 193 AA;
 Query Match 100.0%; Score 193; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 9e-187;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MOSLMAQPLLIAGLLLATPAQAHLLKPSQSFSDNCEGKDPVIRSLTLEDDPIV 60
 DB 1 MOSLMAQPLLIAGLLLATPAQAHLLKPSQSFSDNCEGKDPVIRSLTLEDDPIV 60
 QY 61 PGNVTLVVGVSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYGCTFEHFCVDLMDLIP 120
 DB 61 PGNVTLVVGVSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYGCTFEHFCVDLMDLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLEPLSWLTTCNYRIESVLSGGR 180
 DB 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLEPLSWLTTCNYRIESVLSGGR 180
 QY 181 LGCIKIAASLKI 193
 DB 181 LGCIKIAASLKI 193

RESULT 5
 AAB31928
 ID AAB31928 standard; protein; 193 AA.
 AC AAB31928;
 DT 15-MAY-2001 (first entry)
 DE Amino acid sequence of the human ganglioside GM2 activator protein.
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 FN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INNR) BIOMERIEUX STELHYS.
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI: 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX Disclosure; Fig 1; 209pp; French.
 XX The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX Sequence 193 AA;
 QY Query Match 100.0%; Score 193; DB 4; Length 193;
 Db Best Local Similarity 100.0%; Pred. No. 9e-187;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIALGLLLATPAQHLKKPSQLSSFSWDCDEGKDPVIRSLTLEPDPITV 60
 Db 1 MQSLMQAPLLIALGLLLATPAQHLKKPSQLSSFSWDCDEGKDPVIRSLTLEPDPITV 60
 QY 61 PGNVTLSSVWGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHEFCDVLDMLIP 120
 Db 61 PGNVTLSSVWGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHEFCDVLDMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLPLSWLTGNYRIESVLSGSKR 180
 Db 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLPLSWLTGNYRIESVLSGSKR 180
 QY 181 LGCIKIAASLKG 193
 Db 181 LGCIKIAASLKG 193
 RESULT 6
 AAB31903
 ID AAB31903 standard; protein; 193 AA.
 XX AAB31903;
 XX Amino acid sequence of a human protein.
 XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX Homo sapiens.
 XX WO200105422-A2.
 XX 25-JAN-2001.
 XX 17-JUL-2000; 2000WO-FR002057.
 XX 15-JUL-1999; 99FR-00009372.
 XX (INNR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX Claim 1; Page 164; 209pp; French.
 XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX Sequence 193 AA;
 QY Query Match 100.0%; Score 193; DB 4; Length 193;
 Db Best Local Similarity 100.0%; Pred. No. 9e-187;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIALGLLLATPAQHLKKPSQLSSFSWDCDEGKDPVIRSLTLEPDPITV 60
 Db 1 MQSLMQAPLLIALGLLLATPAQHLKKPSQLSSFSWDCDEGKDPVIRSLTLEPDPITV 60
 QY 61 PGNVTLSSVWGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHEFCDVLDMLIP 120
 Db 61 PGNVTLSSVWGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHEFCDVLDMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLPLSWLTGNYRIESVLSGSKR 180
 Db 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLPLSWLTGNYRIESVLSGSKR 180
 QY 181 LGCIKIAASLKG 193
 Db 181 LGCIKIAASLKG 193
 RESULT 7
 AAB31345
 ID AAB31345 standard; protein; 193 AA.
 XX AAB31345;
 XX 15-NOV-2002 (first entry)
 XX Human GM2 activator protein.
 KW Human; GM2 activator protein; ganglioside; platelet activating factor;
 KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
 KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
 KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
 KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
 KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
 KW anti-inflammatory; antiaesthetic; antiarthritic; antirheumatic;
 KW antiallergic; hepatotropic; nephrotropic; immunosuppressive;
 KW tranquilliser.
 XX Homo sapiens.
 XX OS
 XX US6423680-B1.

XX 23-JUL-2002.
 XX 30-OCT-1998; 98US-00183841.
 XX 30-OCT-1998; 98US-00183841.
 XX (HSCR-) HSC RES & DEV LP.
 XX Rigat B, Reynaud D, Mahuran D;
 PI WPI; 2002-664636/71.
 XX Composition useful for treating inflammatory conditions e.g. asthma
 PT comprises GM2 activator protein or GM2 activator peptide in combination
 PT with a carrier.
 XX Claim 3; Fig 1; 11pp; English.
 XX The present invention relates to a composition comprising GM2
 CC (ganglioside) activator protein or a GM2 activator peptide derived from
 CC the GM2 activator protein in combination with a carrier. The composition
 CC comprises the protein or peptide in an amount of 1-100 mg. The GM2
 CC activator protein is capable of inhibiting platelet activating factor
 CC (PAF). The composition of the invention is useful for treating
 CC inflammatory disorders e.g. inflammatory bowel disease, asthma,
 CC autoimmune disease (such as lupus), hypersensitivity infection,
 CC rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
 CC gout and tissue-specific conditions (e.g. glomerulonephritis and
 CC hepatitis). The composition is capable of inhibiting platelet activating
 CC factor, is non-toxic, is efficacious and presents less severe side
 CC effects, including redness, swelling, pain and polymorphonuclear
 CC leukocyte accumulation at the inflammatory site and other associated
 CC cellular responses. The present sequence represents human GM2 activator
 CC protein
 XX Sequence 193 AA;
 SQ
 Query Match 100.0%; Score 193; DB 5; Length 193;
 Best Local Similarity 100.0%; Pred. No. 9e-187;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIATGALLATPAQAHLKPSQSSFSWDCDEGKOPAVIRSLTLEPDPPIV 60
 Db 1 MQSLMQAPLLIATGALLATPAQAHLKPSQSSFSWDCDEGKOPAVIRSLTLEPDPPIV 60
 QY 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTEFHFCDVLDMLIP 120
 Db 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTEFHFCDVLDMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNYRIEVSLSGKR 180
 Db 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNYRIEVSLSGKR 180
 QY 181 LGCIKIAASLKG 193
 Db 181 LGCIKIAASLKG 193
 RESULT 8
 ID ABP65212 standard; protein; 193 AA.
 XX
 AC ABP65212;
 XX
 DT 12-NOV-2002 (first entry)
 XX
 DE Hypoxia-regulated protein #86.
 XX
 KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;
 KW antiinflammatory; vulnery; gynecological; ophthalmological; vaccine;
 KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
 KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;

presclapmsia; atherosclerosis; inflammatory condition; wound healing;
 inflammation; erythropoiesis; hair loss; human.
 Homo sapiens.
 WO200246465-A2.
 13-JUN-2002.
 10-DEC-2001; 2001WO-GB005458.
 08-DEC-2000; 2000GB-00030076.
 08-FEB-2001; 2001GB-00003156.
 25-OCT-2001; 2001GB-00025666.
 (OXFO-) OXFORD BIOMEDICA UK LTD.
 White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;
 PI Rayner WN;
 XX WPI; 2002-627238/67.
 XX Identifying a gene involved in disease for treating hypoxia-regulated
 PT conditions, comprises comparing the transcriptome/proteome of two cell
 PT types under different conditions and identifying a differentially
 PT regulated gene.
 XX Claim 35; Page 397; 538pp; English.
 XX The present invention relates to methods for identifying genes and
 CC proteins that are implicated in a specific disease or physiological
 CC condition. The method comprises comparing the transcriptome/proteome of a
 CC specialised cell type implicated in a disease or condition with that of a
 CC second specialised cell type, under two experimental conditions, and
 CC identifying a gene that is differentially regulated in the two
 CC specialised cell types under experimental conditions. ABV7873-ABV78116
 CC and ABP65061-ABP65257 were identified using the methods of the invention.
 CC The coding sequences and proteins are useful for treating a disease in a
 CC patient, for manufacture of a medicament for treating hypoxia-regulated
 CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
 CC biological response to hypoxia conditions, or hypoxic-associated
 CC pathology in a patient. The coding sequences and proteins are also useful
 CC for monitoring the therapeutic treatment of a disease or physiological
 CC condition, such as cancer, ischaemic conditions, reperfusion injury,
 CC retinopathy, neonatal stress, presclapmsia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss
 XX Sequence 193 AA;
 SQ
 Query Match 100.0%; Score 193; DB 5; Length 193;
 Best Local Similarity 100.0%; Pred. No. 9e-187;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIATGALLATPAQAHLKPSQSSFSWDCDEGKOPAVIRSLTLEPDPPIV 60
 Db 1 MQSLMQAPLLIATGALLATPAQAHLKPSQSSFSWDCDEGKOPAVIRSLTLEPDPPIV 60
 QY 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTEFHFCDVLDMLIP 120
 Db 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTEFHFCDVLDMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNYRIEVSLSGKR 180
 Db 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNYRIEVSLSGKR 180
 QY 181 LGCIKIAASLKG 193
 Db 181 LGCIKIAASLKG 193
 RESULT 9
 ADN95858
 ID ADN95858 standard; protein; 193 AA.

XX AC ADN95858;
 XX DT 01-JUL-2004 (first entry)
 XX DE Human BEC/LEC-related protein sequence SeqID782.
 XX KW growth; differentiation; blood endothelial cell; BEC;
 KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
 KW lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytosolic;
 KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
 KW inflammatory disease; cancer metastasis; lymphatic system; human.
 XX OS Homo sapiens.
 XX PN WO2003080640-A1.
 XX PD 02-OCT-2003.
 XX PF 07-MAR-2003; 2003WO-US006900.
 XX PR 07-MAR-2002; 2002US-0363019P.
 XX PA (LUDW-) LUDWIG INST CANCER RES.
 XX PA (LICN) LICENTIA LTD.
 XX PI Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
 XX DR WPI; 2003-876899/81.
 XX DR N-PSDB; ADN95859.
 XX PS Example 1; SEQ ID NO 782; 176pp; English.
 XX CC This invention relates to a method of differentially modulating the
 CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
 CC endothelial cells (LEC) comprising contacting endothelial cells with a
 CC composition comprising an agent that differentially modulates blood or
 CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
 CC identifying a human subject with lymphoedema and with a mutation in at
 CC least one allele of a gene encoding a LEC protein, where the mutation
 CC correlates with lymphoedema in human subjects, and with the proviso that
 CC the LEC protein is not VEGFR-3; and administering to the subject a
 CC composition comprising a lymphatic growth agent selected from VEGF-C or
 CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
 CC the development of compounds with an antiangiogenic, cytosolic,
 CC vasotropic or antiinflammatory activity or for gene therapy. The method
 CC is useful in modulating the growth or differentiation of blood
 CC endothelial cells or lymphatic endothelial cells, in treating hereditary
 CC lymphoedema, in screening for an endothelial cell disorder or
 CC predisposition to the disorder or in monitoring the efficacy or toxicity
 CC of a drug on endothelial cells. The agent is useful in manufacturing a
 CC medicament for the differential modulation of blood vessel endothelial
 CC cell or lymphatic vessel endothelial cell growth or differentiation. The
 CC lymphatic growth agent may also be used in manufacturing a medicament for
 CC the treatment of hereditary lymphoedema resulting from a mutation in a
 CC LEC gene or of other diseases involving the lymphatic vessels, such as
 CC various inflammatory diseases and cancer metastasis via the lymphatic
 CC system. The present sequence is that of a human LEC/BEC differentially
 CC expressed protein which is related to the method of the invention. Note:
 CC This sequence does not appear in the specification but was obtained by
 CC the indexer using the source data given in table 14 of the specification.
 XX SQ Sequence 193 AA;
 Query Match 100.0%; Score 193; DB 7; Length 193;
 Best Local Similarity 100.0%; Pred. No. 9e-187;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIALGLLLATPAQAHLKPKSQSSFSWNCDEGKDPVAVIRSLTLEPDPPIV 60
 Db 1 MQSLMQAPLLIALGLLLATPAQAHLKPKSQSSFSWNCDEGKDPVAVIRSLTLEPDPPIV 60
 QY 61 PGNVTLVVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
 DE 100.0%; Pred. No. 9e-187;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIALGLLLATPAQAHLKPKSQSSFSWNCDEGKDPVAVIRSLTLEPDPPIV 60
 Db 1 MQSLMQAPLLIALGLLLATPAQAHLKPKSQSSFSWNCDEGKDPVAVIRSLTLEPDPPIV 60
 QY 61 PGNVTLVVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120

Db 61 PGNVTLVVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPKFKGTYSLPKSEFVVVDLELPWLTGNYRIESVLSGSKR 180
 Db 121 TGEPCPEPLRTYGLPCHCPKFKGTYSLPKSEFVVVDLELPWLTGNYRIESVLSGSKR 180
 QY 181 LGCIKIAASLKG 193
 Db 181 LGCIKIAASLKG 193
 RESULT 10
 ADN03620
 ID ADN03620 standard; protein; 193 AA.
 XX AC ADN03620;
 XX DT 01-JUL-2004 (first entry)
 XX DE Antipsoriatic protein sequence #7.
 XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.
 XX OS Homo sapiens.
 XX PN WO2004028479-A2.
 XX PD 08-APR-2004.
 XX PF 25-SEP-2003; 2003WO-US030907.
 XX PR 25-SEP-2002; 2002US-0414006P.
 XX PA (GETH) GENENTECH INC.
 XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
 XX DR WPI; 2004-305105/28.
 XX DR N-PSDB; ADN03619.
 XX PT New PRO nucleic acid or polypeptide, useful for preparing a
 XX pharmaceutical composition for diagnosing or treating psoriasis in a
 XX mammal.
 XX PS Claim 9; SEQ ID NO 14; 3069pp; English.
 XX CC The invention relates to novel polynucleotide and polypeptides for
 XX treating psoriasis or a sequence having at least 80% identity to the
 XX above sequences. The nucleic acid is useful for preparing a composition
 XX for diagnosing or treating psoriasis in a mammal. This sequence
 XX corresponds to one of the polypeptides of the invention.
 XX SQ Sequence 193 AA;
 Query Match 100.0%; Score 193; DB 8; Length 193;
 Best Local Similarity 100.0%; Pred. No. 9e-187;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIALGLLLATPAQAHLKPKSQSSFSWNCDEGKDPVAVIRSLTLEPDPPIV 60
 Db 1 MQSLMQAPLLIALGLLLATPAQAHLKPKSQSSFSWNCDEGKDPVAVIRSLTLEPDPPIV 60
 QY 61 PGNVTLVVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
 Db 61 PGNVTLVVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPKFKGTYSLPKSEFVVVDLELPWLTGNYRIESVLSGSKR 180
 Db 121 TGEPCPEPLRTYGLPCHCPKFKGTYSLPKSEFVVVDLELPWLTGNYRIESVLSGSKR 180
 QY 181 LGCIKIAASLKG 193

```
Db      181 LGCIKIAASLKGI 193
|||||
RESULT 11
ID AAB31900 standard; protein; 189 AA.
XX
AC AAB31900;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INNR ) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX
DR WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 161-162; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 189 AA;
Query Match 97.9%; Score 189; DB 4; Length 189;
Best Local Similarity 100.0%; Pred. No. 9.9e-183;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 MQAPLLIALGLLATPAQAHLKKPSQLSFSFSDNCDGKDPVIRSLTLEPPIVPGNV 64
Db 1 MQAPLLIALGLLATPAQAHLKKPSQLSFSFSDNCDGKDPVIRSLTLEPPIVPGNV 60
Qy 65 TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFHFCDVLDMLIPTGEP 124
Db 61 TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFHFCDVLDMLIPTGEP 120
Qy 125 CPEPLRTYGLPCHCFKSGTSLPKSEFVVPDLELPFSLMTTGNRYLIESVLSSSGKRLGCI 184
|||||
Db      121 CPEPLRTYGLPCHCFKSGTSLPKSEFVVPDLELPFSLMTTGNRYLIESVLSSSGKRLGCI 180
Qy      185 KIAASLKGI 193
|||||
Db      181 KIAASLKGI 189
|||||
RESULT 12
ID AAB31898 standard; protein; 178 AA.
XX
AC AAB31898;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INNR ) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX
DR WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 160; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 178 AA;
Query Match 92.2%; Score 178; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.3e-171;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 16 LLATPAQAHLKKPSQLSFSFSDNCDGKDPVIRSLTLEPPIVPGNVTLSSVGSTSV 75
Db 1 LLATPAQAHLKKPSQLSFSFSDNCDGKDPVIRSLTLEPPIVPGNVTLSSVGSTSV 60
Qy 76 LSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFHFCDVLDMLIPTGSPCEPLRTYGLP 135
|||||
```

Db 61 LSSPLKVDLVEKEVAGLWIKIPCTDIYIGCTFEHFCVDLMLPTGEPCEPLRTYGLP 120
 Qy 136 CHCFKEGTYSLPKSEFVVPDLELPMSWLTGNYRIEVSLSGKRLGCIKIAASLKGI 193
 Db 121 CHCFKEGTYSLPKSEFVVPDLELPMSWLTGNYRIEVSLSGKRLGCIKIAASLKGI 178

RESULT 13
 ID ABG00720 standard; protein; 193 AA.

AC ABG00720;
 DT 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #711.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.

OS
 XX WO200175067-A2.
 XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX N-PSDB; AAS64907.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 20; SEQ ID NO 31079; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: the sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 193 AA;

Query Match 90.2%; Score 174; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.6e-167;
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 PAQHLKKPSQLSSFSWNCDEGKDPVIRSLTLEPPIVVGNTLSVVGSTSVPLSSP 79
 Db 20 PAQHLKKPSQLSSFSWNCDEGKDPVIRSLTLEPPIVVGNTLSVVGSTSVPLSSP 79
 Qy 80 LKVDLVEKEVAGLWIKIPCTDIYIGCTFEHFCVDLMLPTGEPCEPLRTYGLPCHCP 139
 Db 80 LKVDLVEKEVAGLWIKIPCTDIYIGCTFEHFCVDLMLPTGEPCEPLRTYGLPCHCP 139
 Qy 140 FKEGTYSLPKSEFVVPDLELPMSWLTGNYRIEVSLSGKRLGCIKIAASLKGI 193
 Db 140 FKEGTYSLPKSEFVVPDLELPMSWLTGNYRIEVSLSGKRLGCIKIAASLKGI 193

RESULT 14
 ADQ17712

ID ADQ17712 standard; protein; 193 AA.

XX ADQ17712;
 XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated protein - SEQ ID 529.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

XX 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlotnik A;

XX WPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.

XX Example 2; SEQ ID NO 529; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC protein of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.

XX Sequence 193 AA;

Query Match 90.2%; Score 174; DB 8; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.6e-167;
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 PAQHLKKPSQLSSFSWNCDEGKDPVIRSLTLEPPIVVGNTLSVVGSTSVPLSSP 79
 Db 20 PAQHLKKPSQLSSFSWNCDEGKDPVIRSLTLEPPIVVGNTLSVVGSTSVPLSSP 79

QY 80 LKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIPTGCEPEPLRTYGLPCHCP 139
Db 80 LKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIPTGCEPEPLRTYGLPCHCP 139
QY 140 FKEGTYSLPKSEFVVDLELPSPWLTGNYRIESVLSGKRLGCIKIAASLKG 193
Db 140 FKEGTYSLPKSEFVVDLELPSPWLTGNYRIESVLSGKRLGCIKIAASLKG 193

RESULT 15
ABG31346

ID ABG31346 standard; protein; 178 AA.

XX AC ABG31346;

XX DT 15-NOV-2002 (first entry)

XX DE Non-glycosylated human GM2 activator protein containing His6-tag.

XX KW Human; GM2 activator protein; ganglioside; platelet activating factor;
KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
KW antiinflammatory; antiasthmatic; antiarthritic; antirheumatic;
KW antiallergic; hepatotropic; nephrotropic; immunosuppressive;
KW tranquilizer.

XX OS Homo sapiens.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Binding-site 5..10

FT FT /label= His6 tag
FT FT /note= "Nickel chelating region used for purification of
FT FT the protein"

FT FT 11..16

FT FT /label= Factor_X_cleavage_site

XX US6423680-B1.

XX PD 23-JUL-2002.

XX PF 30-OCT-1998; 98US-00183841.

XX PR 30-OCT-1998; 98US-00183841.

XX PA (HSCR-) HSC RES & DEV LP.

XX PI Rigat B, Reynaud D, Mahuran D;

XX DR WPI; 2002-664636/71.

XX Composition useful for treating inflammatory conditions e.g. asthma
XX comprises GM2 activator protein or GM2 activator peptide in combination
XX with a carrier.

XX Example 1; Fig 2; lpp; English.

XX The present invention relates to a composition comprising GM2
XX (ganglioside) activator protein or a GM2 activator peptide derived from
XX the GM2 activator protein in combination with a carrier. The composition
XX comprises the protein or peptide in an amount of 1-100 mg. The GM2
XX activator protein is capable of inhibiting platelet activating factor
XX (PAF). The composition of the invention is useful for treating
XX inflammatory disorders e.g. inflammatory bowel disease, asthma,
XX autoimmune disease (such as lupus), hypersensitivity infection,
XX rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
XX gout and tissue-specific conditions (e.g. glomerulonephritis and
XX hepatitis). The composition is capable of inhibiting platelet activating
XX factor, is non-toxic, is efficacious and presents less severe side
XX effects, including redness, swelling, pain and polymorphonuclear

CC leukocyte accumulation at the inflammatory site and other associated
CC cellular responses. The present sequence represents a non-glycosylated
CC human GM2 activator protein prepared using a His6-tag bacterial
CC expression system

XX SQ Sequence 178 AA;

Query Match 83.9%; Score 162; DB 5; Length 178;

Best Local Similarity 100.0%; Pred. No. 2.1e-155;

Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SSFSDNDCDEGKDPVIRSLTLEPDPVIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVA 91

Db 17 SSFSDNDCDEGKDPVIRSLTLEPDPVIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVA 76

QY 92 GLWIKIPCTDIYIGSCTFEHFCVDLMDLIPTGCEPEPLRTYGLPCHCPKEGTYSLPKSE 151

Db 77 GLWIKIPCTDIYIGSCTFEHFCVDLMDLIPTGCEPEPLRTYGLPCHCPKEGTYSLPKSE 136

QY 152 FVVPDLELPSPWLTGNYRIESVLSGKRLGCIKIAASLKG 193

Db 137 FVVPDLELPSPWLTGNYRIESVLSGKRLGCIKIAASLKG 178

Search completed: November 17, 2004, 09:12:37

Job time : 33.2046 secs

1115 Page Blank (copy)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 10:19:31 ; Search time 7.70454 Seconds
(without alignments)
1661.276 Million cell updates/sec

Title: US-10-030-937-8

Perfect score: 193

Sequence: 1 MQSLMQAPLLIALGILLATP.....LSSGKRLGCIKIAASLKG1 193

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTCUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	193	100.0	193	4	US-09-183-841-1
2	162	83.9	178	4	US-09-183-841-2
3	9	4.7	262	3	US-08-961-083-98
4	9	4.7	262	4	US-09-536-784-98
5	9	4.7	270	3	US-08-961-083-206
6	9	4.7	270	4	US-09-536-784-206
7	9	4.7	291	4	US-09-583-110-4201
8	8	4.1	444	4	US-09-252-991A-20496
9	8	4.1	563	4	US-09-252-991A-23790
10	8	4.1	769	3	US-09-320-878-12
11	8	4.1	769	4	US-09-141-908-10
12	8	4.1	769	4	US-09-657-440-12
13	8	4.1	809	3	US-09-105-537-24
14	8	4.1	3782	3	US-09-105-537-4
15	7	3.6	85	4	US-09-270-767-32129
16	7	3.6	85	4	US-09-270-767-47346
17	7	3.6	96	3	US-08-936-165A-401
18	7	3.6	103	4	US-09-270-767-33345
19	7	3.6	103	4	US-09-270-767-48562
20	7	3.6	115	4	US-09-746-801A-47
21	7	3.6	153	3	US-09-199-637A-213
22	7	3.6	170	4	US-09-205-258-1011
23	7	3.6	267	4	US-09-489-039A-12889
24	7	3.6	269	4	US-09-543-681A-6475
25	7	3.6	285	1	US-08-149-809-24
26	7	3.6	287	4	US-09-540-236-2879
27	7	3.6	315	4	US-09-252-991A-31850

28	7	3.6	324	4	US-09-489-039A-7803	Sequence 7803, Ap
29	7	3.6	325	2	US-08-828-242-4	Sequence 4, Appli
30	7	3.6	325	3	US-09-206-499-4	Sequence 4, Appli
31	7	3.6	331	2	US-08-828-242-3	Sequence 3, Appli
32	7	3.6	331	2	US-08-910-927B-5	Sequence 5, Appli
33	7	3.6	331	3	US-09-206-499-3	Sequence 3, Appli
34	7	3.6	331	3	US-09-270-270-5	Sequence 5, Appli
35	7	3.6	331	4	US-09-961-403-11	Sequence 11, Appli
36	7	3.6	355	4	US-09-580-929-5	Sequence 5, Appli
37	7	3.6	364	4	US-09-489-039A-8053	Sequence 1008, Ap
38	7	3.6	369	4	US-09-489-039A-8053	Sequence 8053, Ap
39	7	3.6	370	4	US-09-543-681A-4353	Sequence 4353, Ap
40	7	3.6	381	3	US-09-257-580-2	Sequence 2, Appli
41	7	3.6	492	4	US-09-252-991A-32203	Sequence 32203, A
42	7	3.6	536	3	US-09-188-930-185	Sequence 185, App
43	7	3.6	536	4	US-09-312-283C-185	Sequence 185, App
44	7	3.6	590	4	US-09-312-283C-409	Sequence 409, App
45	7	3.6	605	4	US-09-252-991A-24349	Sequence 24349, A

ALIGNMENTS

RESULT 1

US-09-183-841-1
; Sequence 1, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183.841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (33)..(55)
; FEATURE:
; OTHER INFORMATION: residues 56-63 are included in a further precursor
; OTHER INFORMATION: form of the protein
US-09-183-841-1

Query Match	100.0%	Score 193	DB 4	Length 193
Best Local Similarity	100.0%	Pred. No. 8.1e-182	Mismatches 0	Indels 0
Matches 193	Conservative 0			Gaps 0
Qy	1	MQSLMQAPLLIALGILLATPAQAHLKPKSOLSSFSWDCDEGKOPAVIRSLTLEPDP1V 60		
Db	1	MQSLMQAPLLIALGILLATPAQAHLKPKSOLSSFSWDCDEGKOPAVIRSLTLEPDP1V 60		
Qy	61	PGNVTLSVVGSTVPLSSPLKVDLVLEKEVAGLMIKIPCTDIYIGSCTFEHFCVDLMDLIP 120		
Db	61	PGNVTLSVVGSTVPLSSPLKVDLVLEKEVAGLMIKIPCTDIYIGSCTFEHFCVDLMDLIP 120		
Qy	121	TGEPCEPELRTYGLPCHCPKKEGTYSLPKSFVVPDLELPMSLTTGNVRIEVLSSSGKR 180		
Db	121	TGEPCEPELRTYGLPCHCPKKEGTYSLPKSFVVPDLELPMSLTTGNVRIEVLSSSGKR 180		
Qy	181	LGCIKIAASLKG1 193		
Db	181	LGCIKIAASLKG1 193		

RESULT 2

US-09-183-841-2
; Sequence 2, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:

; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: His tag at residues 1 to 17
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: amino acid
; OTHER INFORMATION: sequence of GM2 protein using His6 tag
US-09-183-841-2

Query Match 83.9%; Score 162; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 2.3e-151;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 32 SFSWNCDEGKPAVIRSLTLPDPVIVFGNVTLSVVGSTVPLSSPLKVDLVLEKVA 91
Db 17 SFSWNCDEGKPAVIRSLTLPDPVIVFGNVTLSVVGSTVPLSSPLKVDLVLEKVA 76
QY 92 GLWIKIPCTDIYIGSTFEHFCVLDLMLPTGEPCEPLRTYGLPCHCPFKEGTYSLPKSE 151
Db 77 GLWIKIPCTDIYIGSTFEHFCVLDLMLPTGEPCEPLRTYGLPCHCPFKEGTYSLPKSE 136
QY 152 FVVPDLPLSWLTGNYRIEVSLSGKRLGCIKIAASLKI 193
Db 137 FVVPDLPLSWLTGNYRIEVSLSGKRLGCIKIAASLKI 178

RESULT 3
US-08-961-083-98
; Sequence 98, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brooks, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 98:

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-98
Query Match 4.7%; Score 9; DB 3; Length 262;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 66 LSVVGSTSV 74
Db 155 LSVVGSTSV 163

RESULT 4
US-09-536-784-98
; Sequence 98, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 98:
US-09-536-784-98

Query Match 4.7%; Score 9; DB 4; Length 262;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 66 LSVVGSTSV 74
Db 155 LSVVGSTSV 163
RESULT 5
US-08-961-083-206
; Sequence 206, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.

;; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

;; NUMBER OF SEQUENCES: 452

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Human Genome Sciences, Inc.

;; STREET: 9410 Key West Avenue

;; CITY: Rockville

;; STATE: Maryland

;; COUNTRY: USA

;; ZIP: 20850

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

;; COMPUTER: HP Vectra 486/33

;; OPERATING SYSTEM: MSDOS version 6.2

;; SOFTWARE: ASCII Text

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/961,083

;; FILING DATE:

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER:

;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Brookes, A. Anders

;; REGISTRATION NUMBER: 36,373

;; REFERENCE/DOCKET NUMBER: PB340P2

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (301) 309-8504

;; TELEFAX: (301) 309-8512

;; INFORMATION FOR SEQ ID NO: 206:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 270 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

;; US-08-961-083-206

Query Match 4.7%; Score 9; DB 3; Length 270;

Best Local Similarity 100.0%; Pred. No. 0.9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LSVVGSTSV 74

Db 163 LSVVGSTSV 171

RESULT 6

US-09-536-784-206

Sequence 206, Application US/09536784

Patent No. 6573082

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/536,784

FILING DATE: 30-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: OCT-30-1997

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Michelle S. Marks

;; REGISTRATION NUMBER: 41,971

;; REFERENCE/DOCKET NUMBER: PB340P3

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (301) 309-8504

;; TELEFAX: (301) 309-8512

;; INFORMATION FOR SEQ ID NO: 206:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 270 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

;; SEQUENCE DESCRIPTION: SEQ ID NO: 206:

US-09-536-784-206

Query Match 4.7%; Score 9; DB 4; Length 270;

Best Local Similarity 100.0%; Pred. No. 0.9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LSVVGSTSV 74

Db 163 LSVVGSTSV 171

RESULT 7

US-09-583-110-4201

Sequence 4201, Application US/09583110

Patent No. 6699703

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al.

TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus

FILE REFERENCE: PATH00-07A

CURRENT APPLICATION NUMBER: US/09/583,110

CURRENT FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/107,433

PRIOR FILING DATE: 1998-06-30

PRIOR APPLICATION NUMBER: US 60/085,131

PRIOR FILING DATE: 1998-05-12

PRIOR APPLICATION NUMBER: US 60/051,553

PRIOR FILING DATE: 1997-07-02

NUMBER OF SEQ ID NOS: 5322

SEQ ID NO 4201

LENGTH: 291

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-583-110-4201

Query Match 4.7%; Score 9; DB 4; Length 291;

Best Local Similarity 100.0%; Pred. No. 0.96;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LSVVGSTSV 74

Db 184 LSVVGSTSV 192

RESULT 8

US-09-252-991A-20496

Sequence 20496, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

```

1  APPLICANT: LANG, LI
2  TITLE OF INVENTION: RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE
3  FILE REFERENCE: 300622002120
4  CURRENT APPLICATION NUMBER: US/09/320,878A
5  CURRENT FILING DATE: 1999-05-27
6  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
7  EARLIER FILING DATE: 1998-08-28
8  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
9  EARLIER FILING DATE: 1998-05-06
10 EARLIER APPLICATION NUMBER: CIP OF 08/846,247
11 EARLIER FILING DATE: 1997-04-30
12 EARLIER APPLICATION NUMBER: 60/119,139

```

; APPLICANT: ASHLEY, Gary
 ; APPLICANT: BETLACH, Melanie C.
 ; APPLICANT: BETLACH, Mary C.
 ; APPLICANT: McDANIEL, Robert
 ; APPLICANT: TANG, Li
 ; TITLE OF INVENTION: RECOMBINANT

```
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-657-440-12

Query Match          4.1%; Score 8; DB 4; Length 769;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GLLLATPA 21
Db 285 GLLLATPA 292

RESULT 13
US-09-105-537-24
; Sequence 24, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 809
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-24

Query Match          4.1%; Score 8; DB 3; Length 809;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GLLLATPA 21
Db 344 GLLLATPA 351

RESULT 14
US-09-105-537-4
; Sequence 4, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3782
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae

US-09-105-537-4

Query Match          4.1%; Score 8; DB 3; Length 3782;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GLLLATPA 21
Db 1409 GLLLATPA 1416

RESULT 15
US-09-270-767-32129
; Sequence 32129, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32129
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-32129

Query Match          3.6%; Score 7; DB 4; Length 85;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 LELPSWL 163
Db 53 LELPSWL 59

Search completed: November 17, 2004, 10:53:15
Job time : 8.70454 secs
```

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 10:45:47 ; Search time 25.07 Seconds
(without alignments)
2726.237 Million cell updates/sec

Title: US-10-030-937-8

Perfect score: 193

Sequence: 1 MQSLMQAPLLIALGILLATP.....LSSSGKRLGCIKIAASLKGI 193

Scoring table: OIIGO

Gapop 60.0 , Gapext 60.0

Searched: 1570615 seqs, 354127592 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pcp.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pcp.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193	100.0	193	14	US-10-170-385-389
2	61	31.6	61	9	US-09-864-761-34809
3	11	5.7	11	9	US-09-791-378-579
4	11	5.7	11	11	US-09-791-377-579
5	9	4.7	262	9	US-09-765-272-98
6	9	4.7	270	9	US-09-765-272-206
7	8	4.1	8	9	US-09-791-378-578
8	4.1	4.1	100	11	US-09-791-377-578
9	8	4.1	100	15	US-10-424-599-228216
10	8	4.1	110	17	US-10-425-115-305353
11	8	4.1	170	10	US-09-804-014A-37
12	8	4.1	323	17	US-10-739-930-5722
13	8	4.1	466	16	US-10-437-963-119990

14	8	4.1	574	16	US-10-324-367-36	Sequence 36, Appl
15	8	4.1	589	15	US-10-424-599-245422	Sequence 245422,
16	8	4.1	769	10	US-09-793-708-12	Sequence 12, Appl
17	8	4.1	769	14	US-10-201-365-10	Sequence 10, Appl
18	8	4.1	769	14	US-10-160-539-12	Sequence 12, Appl
19	8	4.1	809	9	US-09-861-289-24	Sequence 24, Appl
20	8	4.1	809	9	US-09-860-846-24	Sequence 24, Appl
21	8	4.1	809	10	US-09-988-384B-24	Sequence 24, Appl
22	8	4.1	809	10	US-09-836-821-24	Sequence 24, Appl
23	8	4.1	809	14	US-10-271-889-24	Sequence 24, Appl
24	8	4.1	809	16	US-10-398-605-24	Sequence 24, Appl
25	8	4.1	1308	16	US-10-437-963-186215	Sequence 186215,
26	8	4.1	3782	9	US-09-861-289-4	Sequence 4, Appl
27	8	4.1	3782	9	US-09-860-846-4	Sequence 4, Appl
28	8	4.1	3782	10	US-09-988-384B-4	Sequence 4, Appl
29	8	4.1	3782	10	US-09-836-821-4	Sequence 4, Appl
30	8	4.1	3782	14	US-10-271-889-47	Sequence 47, Appl
31	8	4.1	3782	16	US-10-398-605-4	Sequence 4, Appl
32	7	3.6	39	11	US-09-833-245-1222	Sequence 1222, Ap
33	7	3.6	39	11	US-09-833-245-1224	Sequence 1224, Ap
34	7	3.6	53	15	US-10-424-599-268608	Sequence 268608,
35	7	3.6	60	15	US-10-424-599-208921	Sequence 208921,
36	7	3.6	60	15	US-10-424-599-241819	Sequence 241819,
37	7	3.6	73	14	US-10-106-698-6961	Sequence 6961, Ap
38	7	3.6	83	15	US-10-425-114-61932	Sequence 61932, A
39	7	3.6	84	17	US-10-425-115-202828	Sequence 202828,
40	7	3.6	87	14	US-10-029-386-31246	Sequence 31246, A
41	7	3.6	89	16	US-10-437-963-143338	Sequence 143338,
42	7	3.6	92	16	US-10-437-963-152413	Sequence 152413,
43	7	3.6	93	17	US-10-425-115-249956	Sequence 249956,
44	7	3.6	96	9	US-09-939-980-401	Sequence 401, App
45	7	3.6	97	15	US-10-424-599-269126	Sequence 269126,

ALIGNMENTS

RESULT 1

US-10-170-385-389
; Sequence 389, Application US/10170385
; Publication No. US20030203372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 389
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-170-385-389

Query Match 100.0%; Score 193; DB 14; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.5e-177;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIALGILLATPQAHLKPKFSQSSFSWNCDEGKDPVIRSLTEPDIIV 60

```

; OTHER INFORMATION: MAP TO AC011342.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99
; OTHER INFORMATION: EST HUMAN HIT: BEL8286.1, EVALUE 9.00e-34
; OTHER INFORMATION: SWISSPROT HIT: P17900, EVALUE 1.00e-34
US-09-864-761-34809

Query Match 31.6%; Score 61; DB 9; Length 61;
Best Local Similarity 100.0%; Pred. No. 9e-51;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 VDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIPGTGCPPEPLRTYGLPCHCPEFK 141
Db 1 VDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIPGTGCPPEPLRTYGLPCHCPEFK 60

Qy 142 E 142
Db 61 E 61

RESULT 3
US-09-791-378-579
; Sequence 579, Application US/09791378
; Patent No. US20020142303A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 9195-061-999
; CURRENT APPLICATION NUMBER: US/09/791,378
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 579
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-378-579

Query Match 5.7%; Score 11; DB 9; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 TYGLPCHCPEFK 141
Db 1 TYGLPCHCPEFK 11

RESULT 4
US-09-791-377-579
; Sequence 579, Application US/09791377
; Publication No. US20040110938A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 9195-060-999
; CURRENT APPLICATION NUMBER: US/09/791,377
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34809
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; OTHER INFORMATION: MAP TO AC011342.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99
; OTHER INFORMATION: EST HUMAN HIT: BEL8286.1, EVALUE 9.00e-34
; OTHER INFORMATION: SWISSPROT HIT: P17900, EVALUE 1.00e-34
US-09-864-761-34809

Query Match 31.6%; Score 61; DB 9; Length 61;
Best Local Similarity 100.0%; Pred. No. 9e-51;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 VDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIPGTGCPPEPLRTYGLPCHCPEFK 141
Db 1 VDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIPGTGCPPEPLRTYGLPCHCPEFK 60

Qy 142 E 142
Db 61 E 61

RESULT 3
US-09-791-378-579
; Sequence 579, Application US/09791378
; Patent No. US20020142303A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 9195-061-999
; CURRENT APPLICATION NUMBER: US/09/791,378
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 579
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-378-579

Query Match 5.7%; Score 11; DB 9; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 TYGLPCHCPEFK 141
Db 1 TYGLPCHCPEFK 11

RESULT 4
US-09-791-377-579
; Sequence 579, Application US/09791377
; Publication No. US20040110938A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 9195-060-999
; CURRENT APPLICATION NUMBER: US/09/791,377
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34809
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; SEQ ID NO 579
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-377-579
    Query Match          5.7%; Score 11; DB 11; Length 11;
    Best Local Similarity 100.0%; Pred. No. 0.003;
    Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPEK 141
    |||||
Db 1 TYGLPCHCPEK 11

RESULT 5
US-09-765-272-98
; Sequence 98, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 98:
US-09-765-272-98
    Query Match          4.7%; Score 9; DB 9; Length 262;
    Best Local Similarity 100.0%; Pred. No. 4.5;
    Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LSVVGSTSV 74
    |||||
Db 155 LSVVGSTSV 163

RESULT 6
US-09-765-272-206
; Sequence 206, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 99:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 99:
US-09-765-272-98
    Query Match          4.7%; Score 9; DB 9; Length 262;
    Best Local Similarity 100.0%; Pred. No. 4.5;
    Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LSVVGSTSV 74
    |||||
Db 155 LSVVGSTSV 163

RESULT 7
US-09-791-378-578
; Sequence 578, Application US/09791378
; Patent No. US20020142303A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 9195-061-999
; CURRENT APPLICATION NUMBER: US/09/791.378
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 578
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-378-578
    Query Match          4.1%; Score 8; DB 9; Length 8;
    Best Local Similarity 100.0%; Pred. No. 1.4e+06;
    Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 206:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 206:
US-09-765-272-206
    Query Match          4.7%; Score 9; DB 9; Length 270;
    Best Local Similarity 100.0%; Pred. No. 4.6;
    Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LSVVGSTSV 74
    |||||
Db 163 LSVVGSTSV 171

RESULT 7
US-09-791-378-578
; Sequence 578, Application US/09791378
; Patent No. US20020142303A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 9195-061-999
; CURRENT APPLICATION NUMBER: US/09/791.378
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 578
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-378-578
    Query Match          4.1%; Score 8; DB 9; Length 8;
    Best Local Similarity 100.0%; Pred. No. 1.4e+06;
    Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 89 EVAGLWIK 96
|||||
Db 1 EVAGLWIK 8

RESULT 8

US-09-791-377-578
; Sequence 578, Application US/09791377
; Publication No. US20040110938A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF
; TITLE OF INVENTION: SCHIZOPHRENIA
; FILE REFERENCE: 9195-060-999
; CURRENT APPLICATION NUMBER: US/09/791.377
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 578
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-377-578

Query Match 4.1%; Score 8; DB 11; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 EVAGLWIK 96
|||||
Db 1 EVAGLWIK 8

RESULT 9

US-10-424-599-228216
; Sequence 228216, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 228216
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_48107C.1.pep
US-10-424-599-228216

Query Match 4.1%; Score 8; DB 15; Length 100;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 STSVPLSS 78
|||||
Db 10 STSVPLSS 17

RESULT 10

US-10-425-115-305353
; Sequence 305353, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 305353
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_41551C.1.pep
US-10-425-115-305353

Query Match 4.1%; Score 8; DB 17; Length 110;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 IRSITLEP 55
|||||
Db 19 IRSITLEP 26

RESULT 11

US-09-804-014A-37
; Sequence 37, Application US/09804014A
; Publication No. US20030064489A1
; GENERAL INFORMATION:
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muraidhara
; APPLICANT: Vernet, Corine
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard
; APPLICANT: Spaderna, Steven
; APPLICANT: Majumder, Kumud
; TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-721 US
; CURRENT APPLICATION NUMBER: US/09/804,014A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/188,316
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/188,277
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/189,139
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/189,140
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/190,401
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/190,231
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 37
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-804-014A-37

Query Match 4.1%; Score 8; DB 10; Length 170;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 LEKEVAGL 93
|||||
Db 107 LEKEVAGL 114

RESULT 12

US-10-739-930-5722


```
; Sequence 5722, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 5722
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C126388_1.p
US-10-739-930-5722

Query Match          4.1%; Score 8; DB 17; Length 323;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      86 LEKEVAGL 93
Db      149 LEKEVAGL 156
|||||

RESULT 13
US-10-437-963-119990
; Sequence 119990, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 119990
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_23152C.1.p
US-10-437-963-119990

Query Match          4.1%; Score 8; DB 16; Length 466;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      86 LEKEVAGL 93
Db      337 LEKEVAGL 344
|||||

RESULT 14
US-10-324-967-36
; Sequence 36, Application US/10324967
; Publication No. US20040122212A1
; GENERAL INFORMATION:
; APPLICANT: Cosson, Pierre
; APPLICANT: Kohler, Thilo
; APPLICANT: Benghezal, Mohammed
; APPLICANT: Marchetti, Anna
; APPLICANT: van Delden, Christian
```

```
; TITLE OF INVENTION: VIRULENCE GENES, PROTEINS, AND THEIR USE
; FILE REFERENCE: 25421-502
; CURRENT APPLICATION NUMBER: US/10/324,967
; CURRENT FILING DATE: 2002-12-20
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-324-967-36

Query Match          4.1%; Score 8; DB 16; Length 574;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      11 IALGULLA 18
Db      154 IALGULLA 161
|||||

RESULT 15
US-10-424-599-245422
; Sequence 245422, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 245422
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_63647C.1.p
US-10-424-599-245422

Query Match          4.1%; Score 8; DB 15; Length 589;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      45 PAVIRSLT 52
Db      321 PAVIRSLT 328
|||||

Search completed: November 17, 2004, 12:23:47
Job time : 26.07 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 07:28:04 ; Search time 32.1224 Seconds
(without alignments)
3457.007 Million cell updates/sec

Title: US-10-030-937-8

Perfect score: 193

Sequence: 1 MQLMQAPLLALGALLATP.....LSSSGRLGCIKIPASLKG1 193

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	193	100.0	193	1	SAP3 HUMAN	P17900 homo sapien
2	189	97.9	189	2	QSLBL5	Q61b15 homo sapien
3	189	97.9	189	2	CAA43994	CAA43994 homo sapi
4	100	51.8	103	2	Q14427	Q14427 homo sapien
5	31	16.1	190	2	Q8HXX6	Q8hxx6 macaca fasc
6	26	13.5	47	2	Q81028	Q8i028 macaca fasc
7	19	9.8	197	2	Q6GLN6	Q6gln6 xenopus lae
8	18	9.3	193	1	SAP3 MOUSE	Q60648 mus musculus
9	18	9.3	199	2	Q8IN37	Q6in37 rattus norv
10	18	9.3	199	2	Q8CJH4	Q8cjh4 rattus norv
11	11	5.7	146	2	Q6Q7X4	Q6q7x4 felis silve
12	11	5.7	146	2	AA564351	AA564351 felis sil
13	11	5.7	151	2	Q6Q7X5	Q6q7x5 felis silve
14	11	5.7	151	2	AA564350	AA564350 felis sil
15	10	5.2	721	2	Q8ALY3	Q8aly3 bacteroides
16	9	4.7	20	2	Q9QUW2	Q9quw2 rattus sp.
17	9	4.7	291	2	Q9X4T0	Q9x4t0 streptococc
18	9	4.7	312	2	Q8DN64	Q8dn64 streptococc
19	9	4.7	401	2	Q88VX3	Q88vx3 lactobacill
20	8	4.1	26	2	Q81029	Q8i029 macaca fasc
21	8	4.1	163	2	Q8RUK1	Q8ruki deinococcus
22	8	4.1	184	2	Q6DT75	Q6dt75 arabidopsis
23	8	4.1	188	1	TFT1_PIG	Q97557 sus scrofa
24	8	4.1	257	1	PNK_RHIME	Q92qj0 rhizobium m
25	8	4.1	300	2	Q47212	Q47212 escherichia
26	8	4.1	302	2	Q7P200	Q7p200 chromobacte
27	8	4.1	323	2	Q8SHU6	Q8shj6 arabidopsis
28	8	4.1	349	2	Q9XN28	Q9xn28 salvelinus
29	8	4.1	357	2	Q9PI06	Q9pi06 campylobact
30	8	4.1	365	2	BAC11346	Bac11346 homo sapi
31	8	4.1	377	2	Q6NBZ3	Q6nbz3 rhodopsueto

32	8	4.1	377	2	CAE26127	CAe26127 rhodopseu
33	8	4.1	378	2	Q8HYU6	Q8hyu6 oryctolagus
34	8	4.1	390	1	TFT1_BOVIN	P27628 bos taurus
35	8	4.1	390	1	TFT1_HUMAN	Q9nnx1 homo sapien
36	8	4.1	414	1	CBX6_MOUSE	Q9dby5 mus musculus
37	8	4.1	444	2	Q6MWL1	Q6mw11 neurospora
38	8	4.1	444	2	CAE76287	CAe76287 neurospor
39	8	4.1	447	2	Q8K5N1	Q6k5n1 oryza sativ
40	8	4.1	447	2	BAD22094	BAD22094 oryza sat
41	8	4.1	448	2	Q95DW6	Q95dw6 goodenia ov
42	8	4.1	467	2	Q762L6	Q762l6 klebsiella
43	8	4.1	467	2	BAD03942	BAD03942 klebsiell
44	8	4.1	511	2	Q8RAZ4	Q8raza4 thermomana
45	8	4.1	525	2	Q7VCD6	Q7vcd6 prochloroco

ALIGNMENTS

RESULT 1
SAP3_HUMAN
ID SAP3_HUMAN STANDARD; PRT; 193 AA.
AC P17900; Q14426; Q14428;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate
DE activator protein) (Shingolipid activator protein 3) (SAP-3).
GN Name=GM2A;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RX MEDLINE=91282768; PubMed=2059210;
RA Xie B., McInnes B., Neote K., Lamhonwah A.-M., Mahuran D.;
RT "Isolation and expression of a full-length cDNA encoding the human G-
M2 activator protein.";
RL Biochem. Biophys. Res. Commun. 177:1217-1223(1991).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
human GM2 activator protein.";
RL FEBS Lett. 289:260-264(1991).
[3]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RC TISSUE=Placenta;
RX MEDLINE=92207171; PubMed=1554364;
RA Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
RT "Evidence for two cDNAs encoding human GM2-activator protein.";
RL Biochem. J. 282:807-813(1992).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93052421; PubMed=1427911;
RA Xie B., Kennedy J.L., McInnes B., Auger D., Mahuran D.J.;
RT "Identification of a processed pseudogene related to the functional
gene encoding the GM2 activator protein: localization of the
pseudogene to human chromosome 3 and the functional gene to human
chromosome 5.";
RL Genomics 14:796-798(1992).
[5]
RP SEQUENCE FROM N.A.
RX MEDLINE=99294584; PubMed=10364519;
RA Chen B., Rigat B., Curry C., Mahuran D.J.;
RT "Structure of the GM2A gene: identification of an exon 2 nonsense
mutation and a naturally occurring transcript with an in-frame
deletion of exon 2.";
RL Am. J. Hum. Genet. 65:77-87(1999).
[6]

SEQUENCE FROM N.A.

TISSUE=Uterus;
MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
Strausberg R.L., Fellings E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Browstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Boak S.A., McWeaver P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hsieh S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.W., Green E.D., Bouffard G.G.,
Blakesley R.W., Touchman J.C., Shevchenko Y., Prange C.,
Rodriguez A.C., Grimwood J., Smyth J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smalhus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[7]
SEQUENCE OF 15-193 FROM N.A.
MEDLINE=89325664; PubMed=2753159;
Schroeder M., Klima H., Nakano T., Kwon H., Quintern L.E.,
Gaertner S., Suzuki K., Sandhoff K.;
"Isolation of a cDNA encoding the human GM2 activator protein.";
FEBS Lett. 251:197-200(1989).
[8]
SEQUENCE OF 32-193.
TISSUE=Kidney;
MEDLINE=91006165; PubMed=2209618;
Furst W., Schubert J., Machleidt W., Meyer H.E., Sandhoff K.;
"The complete amino-acid sequences of human ganglioside GM2 activator
protein and cerebroside sulfate activator protein.";
Eur. J. Biochem. 192:709-714(1990).
[9]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
MEDLINE=2045600; PubMed=11090283; DOI=10.1006/jmbi.2000.4225;
Wright C.S., Li S.-C., Rastinejad F.;
"Crystal structure of human GM2-activator protein with a novel beta-
cup topology.";
J. Mol. Biol. 304:411-422(2000).
[10]
VARIANT TSD-AB ARG-138.
MEDLINE=92008638; PubMed=1915858;
Schroeder M., Schnabel D., Suzuki K., Sandhoff K.;
"A mutation in the gene of a glycolipid-binding protein (GM2
activator) that causes GM2-gangliosidosis variant AB.";
FEBS Lett. 290:1-3(1991).
[11]
VARIANT TSD-AB PRO-169.
MEDLINE=94063850; PubMed=8244332;
Schroeder M., Schnabel D., Hurwitz R., Young E., Suzuki K.,
Sandhoff K.;
"Molecular genetics of GM2-gangliosidosis AB variant: a novel mutation
and expression in BHK cells.";
Hum. Genet. 92:437-440(1993).
[12]
VARIANT TSD-AB LYS-88 DEL.
MEDLINE=97055887; PubMed=8900233;
Schepers U., Gombitza G., Lemm T., Hoffmann A., Chabas A., Orand P.,
Sandhoff K.;
"Molecular analysis of a GM2-activator deficiency in two patients with
GM2-gangliosidosis AB variant.";
Am. J. Hum. Genet. 59:1048-1056(1996).
[13]
FUNCTION: Binds gangliosides and stimulates ganglioside GM2
degradation. It stimulates only the breakdown of ganglioside GM2
and glycolipid GA2 by beta-hexosaminidase A. It extracts single
GM2 molecules from membranes and presents them in soluble form to
beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and
CC

```
FT HELIX 111 118
FT TURN 121 122

Query Match 100.0%; Score 193; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.7e-186;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOSLMOAPLLIALGILLATPAQAHLLKPSQSSFSWDCDEKDPVIRSLTLEPDPPIV 60
Db 1 MOSLMOAPLLIALGILLATPAQAHLLKPSQSSFSWDCDEKDPVIRSLTLEPDPPIV 60

QY 61 PGNVTLVVVGSTVPSLSPKVDLVLEKEVAGLWIKIPCTDYIGSTCFHFCDVLDMLIP 120
Db 61 PGNVTLVVVGSTVPSLSPKVDLVLEKEVAGLWIKIPCTDYIGSTCFHFCDVLDMLIP 120

QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNGYRIEVSLSGKRLGCI 180
Db 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNGYRIEVSLSGKRLGCI 180

QY 181 LGCIKIAASLKG 193
Db 181 LGCIKIAASLKG 193

RESULT 2
Q6LBP5 PRELIMINARY; PRT; 189 AA.
AC Q6LBP5;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE GM2 activator protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
RL human GM2 activator protein."
RL FEBS Lett. 289:260-264 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277527; PubMed=8503891;
RA Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA Sandhoff K.;
RT "Over-expression of a functionally active human GM2-activator protein
RT in escherichia coli."
RL Biochem. J. 292:571-576 (1993).
DR EMBL; X62078; CAA43994.1; -.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match 97.9%; Score 189; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.9e-182;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 MQAPLLIALGILLATPAQAHLLKPSQSSFSWDCDEKDPVIRSLTLEPDPPIVPGNV 64
Db 1 MQAPLLIALGILLATPAQAHLLKPSQSSFSWDCDEKDPVIRSLTLEPDPPIVPGNV 60

QY 65 TLSVVGSTVPSLSPKVDLVLEKEVAGLWIKIPCTDYIGSTCFHFCDVLDMLIPTGEP 124
Db 61 TLSVVGSTVPSLSPKVDLVLEKEVAGLWIKIPCTDYIGSTCFHFCDVLDMLIPTGEP 120

QY 125 CPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNGYRIEVSLSGKRLGCI 184
Db 121 CPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNGYRIEVSLSGKRLGCI 180

QY 185 KIAASLKG 193
Db 181 KIAASLKG 189

RESULT 4
Q14427 PRELIMINARY; PRT; 103 AA.
AC Q14427;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GM2-activator protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=92207171; PubMed=1554364;
```

RA Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
 RT "Evidence for two cDNAs encoding human GM2-activator protein.";
 RL Biochem. J. 282:807-813 (1992).
 DR EMBL; X61094; CAA43407.1; -.
 FT NON_TER 1 1
 SQ SEQUENCE 103 AA; 11137 MW; 434BFC334E976F91 CRC64;
 Query Match 51.8%; Score 100; DB 2; Length 103;
 Best Local Similarity 100.0%; Pred. No. 1.2e-92;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 43 KDPVIRSLTLEDDPIVPCNVLTISVVGSTVPLSSPKVDLVLEKVGAGLWIKIPCTDY 102
 DB 1 KDPVIRSLTLEDDPIVPCNVLTISVVGSTVPLSSPKVDLVLEKVGAGLWIKIPCTDY 60
 QY 103 IGSCTFEHFCVDLMDLPTGCPPEPLRTYGLPCHCPFE 142
 DB 61 IGSCTFEHFCVDLMDLPTGCPPEPLRTYGLPCHCPFE 100
 RESULT 5
 Q8HXX6 PRELIMINARY; PRT; 190 AA.
 ID Q8HXX6
 AC Q8HXX6;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Ganglioside GM2 activator.
 GN Name=g2a;
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain cerebellum cortex;
 RA Kusuda J., Osada N., Hida M., Sugano S., Hashimoto K.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB083313; BAC20592.1; -.
 DR HSP; P17900; IG13.
 DR InterPro; IPR003172; El_DerP2_DerF2.
 DR SMART; SM00737; ML; 1
 SQ SEQUENCE 190 AA; 20494 MW; 9F9582BEB75715C3 CRC64;
 Query Match 16.1%; Score 31; DB 2; Length 190;
 Best Local Similarity 100.0%; Pred. No. 1.4e-22;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 125 CPEPLRTYGLPCHCPFKFEGTYSLPKSEFVVP 155
 DB 122 CPEPLRTYGLPCHCPFKFEGTYSLPKSEFVVP 152
 RESULT 6
 Q8I028 PRELIMINARY; PRT; 47 AA.
 ID Q8I028;
 AC Q8I028;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Ganglioside GM2 activator (Fragment).
 GN Name=g2a;
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kusuda J., Osada N., Hashimoto K.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB083332; BAC20643.1; -.

FT NON_TER 1 1
 FT NON_TER 47 47
 SQ SEQUENCE 47 AA; 4883 MW; C5F5537F3A029FFB CRC64;
 Query Match 13.5%; Score 26; DB 2; Length 47;
 Best Local Similarity 100.0%; Pred. No. 4.3e-18;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 33 SPSWDCDEGKDPVIRSLTLEDDPI 58
 DB 3 SPSWDCDEGKDPVIRSLTLEDDPI 28
 RESULT 7
 Q6GLN6 PRELIMINARY; PRT; 197 AA.
 ID Q6GLN6
 AC Q6GLN6;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 MDLINE=22398257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Johnson R.D., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 MDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 initiative.";
 RL Dev. Dyn. 225:384-391 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 RA Klein S.L., Strausberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC074424; AAH74424.1; -.
 DR InterPro; IPR003172; El_DerP2_DerF2.
 DR SMART; SM00737; ML; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 197 AA; 21401 MW; A0909CAD82FD66EB CRC64;
 Query Match 9.8%; Score 19; DB 2; Length 197;
 Best Local Similarity 100.0%; Pred. No. 2e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 125 CPEPLRTYGLPCHCPFKG 143
Db 125 CPEPLRTYGLPCHCPFKG 143

RESULT 8
ID3_SAP3_MOUSE STANDARD; PRT; 193 AA.
AC Q61648; Q61610; Q61819;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate
activator protein) (Shingolipid activator protein 3) (SAP-3).
GN Name=GM2a;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=95229165; PubMed=7713516;
RA Yamanaka S., Johnson O.N., Lyu M.S., Kozak C.A., Proia R.L.;
RT "The mouse gene encoding the GM2 activator protein (Gm2a): cDNA
sequence, expression, and chromosome mapping.";
RL Genomics 24:601-604(1994).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93371367; PubMed=7689829;
RA Bellachiona G., Stirling J.L., Orlacchio A., Beccari T.;
RT "Cloning and sequence analysis of a cDNA clone coding for the mouse
GM2 activator protein.";
RL Biochem. J. 294:227-230(1993).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA;
RX MEDLINE=97224573; PubMed=9060405;
RA Bertoni C., Appolloni M.G., Stirling J.L., Li S.C., Li Y.T.,
RX Orlacchio A., Beccari T.;
RT "Structural organization and expression of the gene for the mouse GM2
activator protein.";
RL Mamm. Genome 8:90-93(1997).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Binds gangliosides and stimulates ganglioside GM2
degradation. It stimulates only the breakdown of ganglioside GM2
and glycolipid GA2 by beta-hexosaminidase A. It extracts single
GM2 molecules from membranes and presents them in soluble form to
beta-hexosaminidase A for cleavage of N-acetyl-D-glucosamine and

```

```

CC conversion to GM3.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney and
CC testis.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U09816; AAA21543.1; -.
CC EMBL; L19526; AAA61929.1; -.
CC EMBL; U34359; AAB06275.1; ALT SEQ.
CC EMBL; U34356; AAB06275.1; JOINED.
CC EMBL; U34357; AAB06275.1; JOINED.
CC EMBL; U34358; AAB06275.1; JOINED.
CC EMBL; BC004651; AAB04651.1; -.
CC HSSP; P17900; 1G13.
CC MGD; MGI:95762; Gm2a.
KW Glycoprotein; Lysosome; Signal; Sphingolipid metabolism.
FT SIGNAL 1 31 By similarity.
FT CHAIN 32 193 Ganglioside GM2 activator.
FT DISULFID 39 183 By similarity.
FT DISULFID 99 106 By similarity.
FT DISULFID 112 138 By similarity.
FT DISULFID 125 136 By similarity.
FT CARBOHYD 151 151 N-linked (GLCNAC...) (Potential).
FT CONFLICT 53 53 I -> T (in Ref.1).
SQ SEQUENCE 193 AA; 20824 MW; 59CC4ABES6FA1FC7 CRC64;
Query Match 9.3%; Score 18; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 2e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 131 TYGLPCHCPFKGTYSLP 148
Db 131 TYGLPCHCPFKGTYSLP 148
RESULT 9
Q6IN37 PRELIMINARY; PRT; 199 AA.
ID Q6IN37
AC Q6IN37;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE GM2 ganglioside activator protein.
GN Name=Gm2a;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

```

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalek U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC072474; AAH72474.1; -;
 DR InterPro: IPR003172; E1_DerP2_DerF2.
 DR SMART: SM00737; ML; 1.
 SQ SEQUENCE 199 AA; 21493 MW; C2B5203FDFABF507 CRC64;

Query Match 9.3%; Score 18; DB 2; Length 199;
 Best Local Similarity 100.0%; Pred. No. 2e-09;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPKPGTYSLP 148
 |||||
 Db 137 TYGLPCHCPKPGTYSLP 154

RESULT 10

ID Q8CUH4 PRELIMINARY; PRT; 199 AA.
 AC Q8CUH4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE GM2 activator protein.
 GN Name=GM2AP;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Miwa N., Okada T., Nakamura S.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB051391; BAC24018.1; -;
 DR HSP; P17900; IG13.
 DR InterPro: IPR003172; E1_DerP2_DerF2.
 DR SMART: SM00737; ML; 1.
 SQ SEQUENCE 199 AA; 21521 MW; D585203FDFABF507 CRC64;

Query Match 9.3%; Score 18; DB 2; Length 199;
 Best Local Similarity 100.0%; Pred. No. 2e-09;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPKPGTYSLP 148
 |||||
 Db 137 TYGLPCHCPKPGTYSLP 154

RESULT 11

ID Q6Q7X4 PRELIMINARY; PRT; 146 AA.
 AC Q6Q7X4;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE GM2 activator protein variant AB (Fragment).
 GN Name=GM2A;
 OS Felis silvestris catus (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 OX NCBI_TaxID=9685;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
 RA Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

RA Dodson A.N., Gentry A.S., Griffin B., Rolms M.D., Baker H.J.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY553654; AAS64351.1; -;
 DR InterPro: IPR003172; E1_DerP2_DerF2.
 DR SMART: SM00737; ML; 1.
 FT NON TER 1
 SQ SEQUENCE 146 AA; 15820 MW; 754E06C0CFCFA74E3 CRC64;

Query Match 5.7%; Score 11; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 0.018;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPK 141
 |||||
 Db 89 TYGLPCHCPK 99

RESULT 12

AAS64351 PRELIMINARY; PRT; 146 AA.
 ID AAS64351
 AC AAS64351;
 DT 25-MAR-2004 (TrEMBLrel. 27, Created)
 DT 25-MAR-2004 (TrEMBLrel. 27, Last sequence update)
 DT 25-MAR-2004 (TrEMBLrel. 27, Last annotation update)
 DE GM2 activator protein variant AB (Fragment).
 GN GM2A.
 OS Felis silvestris catus (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 OX NCBI_TaxID=9685;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
 RA Dodson A.N., Gentry A.S., Griffin B., Rolms M.D., Baker H.J.;
 RT "Mutation of the GM2 activator protein in a feline model of GM2
 RT gangliosidosis";
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY553654; AAS64351.1; -;
 FT NON TER 1
 SQ SEQUENCE 146 AA; 15820 MW; 754E06C0CFCFA74E3 CRC64;

Query Match 5.7%; Score 11; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 0.018;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPK 141
 |||||
 Db 89 TYGLPCHCPK 99

RESULT 13

Q6Q7X5 PRELIMINARY; PRT; 151 AA.
 ID Q6Q7X5
 AC Q6Q7X5;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE GM2 activator protein (Fragment).
 GN Name=GM2A;
 OS Felis silvestris catus (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 OX NCBI_TaxID=9685;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
 RA Dodson A.N., Gentry A.S., Griffin B., Rolms M.D., Baker H.J.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY553653; AAS64350.1; -;
 DR InterPro: IPR003172; E1_DerP2_DerF2.
 DR SMART: SM00737; ML; 1.
 FT NON TER 1
 SQ SEQUENCE 151 AA; 16310 MW; E23EF16CE5844A5F CRC64;

Query Match 5.7%; Score 11; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPEK 141
|||||
Db 89 TYGLPCHCPEK 99

RESULT 14
AAS64350
ID AAS64350 PRELIMINARY; PRT; 151 AA.
AC AAS64350;
DT 25-MAR-2004 (TREMELrel. 27, Created)
DT 25-MAR-2004 (TREMELrel. 27, Last sequence update)
DT 25-MAR-2004 (TREMELrel. 27, Last annotation update)
DE GM2 activator protein (Fragment).
GN GM2A.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
RA Dodson A.N., Gentry A.S., Griffin B., Rolfsma M.D., Baker H.J.;
RT "Mutation of the GM2 activator protein in a feline model of GM2
RT gangliosidosis."
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY553653; AAS64350.1; --
FT NON_TER
SQ SEQUENCE 151 AA; 16310 MW; E23EF16CE5844A5F CRC64;

Query Match 5.7%; Score 11; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPEK 141
|||||
Db 89 TYGLPCHCPEK 99

RESULT 15
Q8A1Y3
ID Q8A1Y3 PRELIMINARY; PRT; 721 AA.
AC Q8A1Y3;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=BT3525;
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."
RL Science 299:2074-2076 (2003).
DR EMBL: AE016941; AAO78631.1; --
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 721 AA; 83162 MW; 9F02F07EAA991182 CRC64;

Query Match 5.2%; Score 10; DB 2; Length 721;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LLIALGLLLA 18

Db 6 LLIALGLLLA 15

Search completed: November 17, 2004, 09:34:48
Job time : 34.1224 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 05:38:38 ; Search time 31.3737 Seconds
(without alignments)
2206.779 Million cell updates/sec

Title: US-10-030-937-8

Perfect score: 1018

Sequence: 1 MQSLMQAPLLIALGLLITATP.....LSSGKRLGCIKIAASLKGI 193

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_23Sep04: *
1: geneseqp1980s: *
2: geneseqp1990s: *
3: geneseqp2000s: *
4: geneseqp2001s: *
5: geneseqp2002s: *
6: geneseqp2003as: *
7: geneseqp2003bs: *
8: geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1018	100.0	193	4 AAB31901	Aab31901 Amino aci
2	1018	100.0	193	4 AAB31902	Aab31902 Amino aci
3	1018	100.0	193	4 AAB31904	Aab31904 Amino aci
4	1018	100.0	193	4 AAB31896	Aab31896 Amino aci
5	1018	100.0	193	4 AAB31928	Aab31928 Amino aci
6	1018	100.0	193	4 AAB31903	Aab31903 Amino aci
7	1018	100.0	193	5 AAB31345	Abg311345 Human GM2
8	1018	100.0	193	5 ABP65212	Abp65212 Hypoxia-r
9	1018	100.0	193	7 AAB31897	Adn5858 Human BEC
10	1018	100.0	193	8 AAB31900	Adn03620 Antipsori
11	1013	99.5	193	4 AAB31898	Abg00720 Novel hum
12	1013	99.5	193	8 AAB31899	Adg17712 Human sof
13	1009	99.1	200	4 AAB31899	Aab31899 Amino aci
14	1005	98.7	193	4 AAB31897	Aab31897 Amino aci
15	1000	98.2	189	4 AAB31900	Aab31900 Amino aci
16	949	93.2	178	4 AAB31898	Aab31898 Amino aci
17	870	85.5	178	5 AAB31346	Abg311346 Non-glyco
18	748.5	73.5	199	2 AAW10656	Aaw10656 Rat GM2 a
19	419	41.2	76	4 AAG00717	Abg00717 Novel hum
20	411	40.4	191	5 AAB31479	Adn41479 Human ova
21	353	34.7	61	4 AAM15082	Aam15082 Peptide #
22	353	34.7	61	4 ABB34073	Abb34073 Peptide #
23	353	34.7	61	4 AAM27530	Aam27530 Peptide #
24	353	34.7	61	4 ABB28897	Abb28897 Peptide #
25	353	34.7	61	4 ABB19511	Abb19511 Protein #

26	353	34.7	61	4 AAM67236	Aam67236 Human bon
27	353	34.7	61	4 AAM54855	Aam54855 Human bra
28	353	34.7	61	4 ABG48902	Abg48902 Human liv
29	353	34.7	61	4 AAM02818	Aam02818 Peptide #
30	353	34.7	61	5 ABG36887	Abg36887 Human pep
31	162.5	16.0	126	4 AAM96319	Aam96319 Human rep
32	122	12.0	152	3 AAG20115	Aag20115 Arabidops
33	122	12.0	152	3 AAG27422	Aag27422 Arabidops
34	122	12.0	156	3 AAG27421	Aag27421 Arabidops
35	122	12.0	161	3 AAG20114	Aag20114 Arabidops
36	116	11.4	131	4 AAG00719	Aag00719 Novel hum
37	115	11.3	152	3 AAG13627	Aag13627 Arabidops
38	95	9.3	277	8 ADL27514	Adl27514 Amino aci
39	91	8.9	160	3 AAG61122	Aag61122 Arabidops
40	91	8.9	160	3 AAG18322	Aag18322 Arabidops
41	90	8.8	456	5 ABB92930	Abb92930 Herbicida
42	89.5	8.8	150	4 AAU47794	Aau47794 Propionib
43	89.5	8.8	150	4 AAU66351	Aau66351 Propionib
44	89.5	8.8	150	6 AEM44313	Abm44313 Propionib
45	89.5	8.8	150	6 AEM62870	Abm62870 Propionib

ALIGNMENTS

RESULT 1
AAB31901
ID AAB31901 standard; protein; 193 AA.

XX AC AAB31901;

XX DT 15-MAY-2001 (first entry)

XX DE Amino acid sequence of a human protein.

XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX OS Homo sapiens.

XX PN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99FR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX DR WPI; 2001-159475/16.

XX PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.

XX PS Claim 1; Page 162-163; 209pp; French.

XX CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and

CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 100.0%; Score 1018; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.9e-104;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIALGLLLATPAQAHLKPKSOLFSSFWDCDEGKDPVIRSLTLEPPIV 60
 DB 1 MQSLMQAPLLIALGLLLATPAQAHLKPKSOLFSSFWDCDEGKDPVIRSLTLEPPIV 60
 QY 61 PGNVTLVSVGSTSVPLSSPLKVDLVEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
 DB 61 PGNVTLVSVGSTSVPLSSPLKVDLVEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
 QY 121 TGEPCEPEPLRTYGLPCHCPKPGCTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180
 DB 121 TGEPCEPEPLRTYGLPCHCPKPGCTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180
 QY 181 LGCIKIAASLKI 193
 DB 181 LGCIKIAASLKI 193

RESULT 2
 AAB31902
 ID AAB31902 standard; protein; 193 AA.
 XX
 AC AAB31902;
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of a human protein.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; Gene therapy.

XX Homo sapiens.
 OS
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.

XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 1; Page 163; 209pp; French.

XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a

CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 100.0%; Score 1018; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.9e-104;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIALGLLLATPAQAHLKPKSOLFSSFWDCDEGKDPVIRSLTLEPPIV 60
 DB 1 MQSLMQAPLLIALGLLLATPAQAHLKPKSOLFSSFWDCDEGKDPVIRSLTLEPPIV 60
 QY 61 PGNVTLVSVGSTSVPLSSPLKVDLVEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
 DB 61 PGNVTLVSVGSTSVPLSSPLKVDLVEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
 QY 121 TGEPCEPEPLRTYGLPCHCPKPGCTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180
 DB 121 TGEPCEPEPLRTYGLPCHCPKPGCTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180
 QY 181 LGCIKIAASLKI 193
 DB 181 LGCIKIAASLKI 193

RESULT 3
 AAB31904
 ID AAB31904 standard; protein; 193 AA.
 XX
 AC AAB31904;
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of a human protein.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX Homo sapiens.
 OS
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.

XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 1; Page 164-165; 209pp; French.

XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses

at least one polypeptide or polynucleotide sequence belonging to the perlecan, precursor of the retinol-binding plasma protein, precursor of the ganglioside GM2 activator, calgranulin B or saposin B protein families. The method is used for detecting, preventing or treating a degenerative, neurological and/or auto-immune disease. The polynucleotides and polypeptides are used for diagnosis, prognosis, prevention and treatment of multiple sclerosis (in its various forms and phases). They may also be useful in cases of e.g. Alzheimer's and Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid polyarthritis and lupus erythematosus, including use as vaccines and in gene therapy (expression of sense or antisense sequences). They can also be used to assess efficacy of potential therapeutic agents, particularly compounds that reduce or inhibit toxicity towards glial cells

Sequence 193 AA;

Query Match 100.0%; Score 1018; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.9e-104;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIAGLLIATPAQAHLKPSQSSFSWDCDEGKDPAVIRSLTLEPDPPIV 60
Db 1 MQSLMQAPLLIAGLLIATPAQAHLKPSQSSFSWDCDEGKDPAVIRSLTLEPDPPIV 60
QY 61 PGNVTLVVGSGTSPVLSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIP 120
Db 61 PGNVTLVVGSGTSPVLSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIP 120
QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSFVVPDLELPWLTGNYRIEVSLSGKR 180
Db 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSFVVPDLELPWLTGNYRIEVSLSGKR 180
QY 181 LGCIKIAASLKG 193
Db 181 LGCIKIAASLKG 193

RESULT 4

AAB31896
ID AAB31896 standard; protein; 193 AA.

XX AC AAB31896;

XX DT 15-MAY-2001 (first entry)

XX DE Amino acid sequence of the human ganglioside GM2 activator protein.

XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
XX KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
XX KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
XX KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
XX KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX OS Homo sapiens.

XX PN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99FR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX DR WPI; 2001-159475/16.

XX DR N-PSDB; AAF54698.

XX FT Detecting, preventing and treating degenerative, neurological and
XX FT autoimmune diseases, particularly multiple sclerosis, using specified
XX FT polypeptides or related nucleic acid or ligand.

XX PS

XX Claim 23; Page 158-159; 209pp; French.

CC The present sequence represents a human polypeptide, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 193 AA;

Query Match 100.0%; Score 1018; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.9e-104;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIAGLLIATPAQAHLKPSQSSFSWDCDEGKDPAVIRSLTLEPDPPIV 60

Db 1 MQSLMQAPLLIAGLLIATPAQAHLKPSQSSFSWDCDEGKDPAVIRSLTLEPDPPIV 60

QY 61 PGNVTLVVGSGTSPVLSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIP 120

Db 61 PGNVTLVVGSGTSPVLSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIP 120

QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSFVVPDLELPWLTGNYRIEVSLSGKR 180

Db 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSFVVPDLELPWLTGNYRIEVSLSGKR 180

QY 181 LGCIKIAASLKG 193

Db 181 LGCIKIAASLKG 193

RESULT 5

AAB31928
ID AAB31928 standard; protein; 193 AA.

XX AC AAB31928;

XX DT 15-MAY-2001 (first entry)

XX DE Amino acid sequence of the human ganglioside GM2 activator protein.

XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
XX KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
XX KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
XX KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
XX KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX OS Homo sapiens.

XX PN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99FR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
XX Disclosure; Fig 1; 209pp; French.
XX
XX The present sequence represents a human polypeptide, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
XX Sequence 193 AA;
SQ
Query Match 100.0%; Score 1018; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.9e-104;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQSIMGAPLIALGLLLATPAQAHKKPSQLSFSFSDNCDEGKDPVIRSLTLEPDPV 60
DB 1 MQSIMGAPLIALGLLLATPAQAHKKPSQLSFSFSDNCDEGKDPVIRSLTLEPDPV 60
QY 61 PGNVTLVSGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDLMLIP 120
DB 61 PGNVTLVSGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDLMLIP 120
QY 121 TGEPCPEPLTYGLPCHCPKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGKR 180
DB 121 TGEPCPEPLTYGLPCHCPKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGKR 180
QY 181 LGCIKTAASLKI 193
DB 181 LGCIKTAASLKI 193
RESULT 6
AAB31903
ID AAB31903 standard; protein; 193 AA.
XX
XX AAB31903;
XX
XX 15-MAY-2001 (first entry)
XX
XX Amino acid sequence of a human protein.
XX
XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
XX Homo sapiens.
OS
XX WO200105422-A2.
FN
XX 25-JAN-2001.
PD
XX 17-JUL-2000; 2000WO-FR002057.
DF
XX 15-JUL-1999; 99FR-00009372.
PR
XX (INNR) BIOMERIEUX STELHYS.
PA

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
PI WPI; 2001-159475/16.
XX
XX Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
XX Claim 1; Page 164; 209pp; French.
XX
XX The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
XX Sequence 193 AA;
SQ
Query Match 100.0%; Score 1018; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.9e-104;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQSIMGAPLIALGLLLATPAQAHKKPSQLSFSFSDNCDEGKDPVIRSLTLEPDPV 60
DB 1 MQSIMGAPLIALGLLLATPAQAHKKPSQLSFSFSDNCDEGKDPVIRSLTLEPDPV 60
QY 61 PGNVTLVSGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDLMLIP 120
DB 61 PGNVTLVSGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDLMLIP 120
QY 121 TGEPCPEPLTYGLPCHCPKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGKR 180
DB 121 TGEPCPEPLTYGLPCHCPKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGKR 180
QY 181 LGCIKTAASLKI 193
DB 181 LGCIKTAASLKI 193
RESULT 7
AAB31345
ID AAB31345 standard; protein; 193 AA.
XX
XX AAB31345;
XX
XX 15-NOV-2002 (first entry)
XX
XX Human GM2 activator protein.
XX
XX Human; GM2 activator protein; ganglioside; platelet activating factor;
KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
KW antiinflammatory; antiasthmatic; antiarthritic; antirheumatic;
KW antiallergic; hepatotropic; nephrotropic; immunosuppressive;
XX
XX Homo sapiens.
OS
XX US6423680-B1.
FN

XX PD 23-JUL-2002.
 XX PF 30-OCT-1998; 98US-00183841.
 XX PR 30-OCT-1998; 98US-00183841.
 XX PA (HSCR-) HSC RES & DEV LP.
 XX PI Rigat B, Reynaud D, Mahuran D;
 XX WPI; 2002-664636/71.
 XX Composition useful for treating inflammatory conditions e.g. asthma
 PT comprises GM2 activator protein or GM2 activator peptide in combination
 PT with a carrier.
 XX Claim 3; Fig 1; 11pp; English.
 XX The present invention relates to a composition comprising GM2
 CC (ganglioside) activator protein or a GM2 activator peptide derived from
 CC the GM2 activator protein in combination with a carrier. The composition
 CC comprises the protein or peptide in an amount of 1-100 mg. The GM2
 CC activator protein is capable of inhibiting platelet activating factor
 CC (PAF). The composition of the invention is useful for treating
 CC inflammatory disorders e.g. inflammatory bowel disease, asthma,
 CC autoimmune disease (such as lupus), hypersensitivity infection,
 CC rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
 CC gout and tissue-specific conditions (e.g. glomerulonephritis and
 CC hepatitis). The composition is capable of inhibiting platelet activating
 CC factor, is non-toxic, is efficacious and presents less severe side
 CC effects, including redness, swelling, pain and polymorphonuclear
 CC leukocyte accumulation at the inflammatory site and other associated
 CC cellular responses. The present sequence represents human GM2 activator
 CC protein
 XX Sequence 193 AA;
 SQ

Query Match 100.0%; Score 1018; DB 5; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.9e-104;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIAGLLLATPAQAHLKPSQSSFSWNCDEGKDPVIRSLTLEPDPV 60
 Db 1 MQSLMQAPLLIAGLLLATPAQAHLKPSQSSFSWNCDEGKDPVIRSLTLEPDPV 60
 QY 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGCTFEHFCVDLMDLIP 120
 Db 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGCTFEHFCVDLMDLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPWLTGNVRIESVLSGSKR 180
 Db 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPWLTGNVRIESVLSGSKR 180
 QY 181 LGCIIKIAASLKG 193
 Db 181 LGCIIKIAASLKG 193

RESULT 8

ABP65212
 ID ABP65212 standard; protein; 193 AA.

XX AC ABP65212;
 XX 12-NOV-2002 (first entry)

XX DE Hypoxia-regulated protein #86.
 KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;
 KW antiinflammatory; vulnary; gynecological; ophthalmological; vaccine;
 KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
 KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;

KW preclapmsia; atherosclerosis; inflammatory condition; wound healing;
 KW inflammation; erythropoiesis; hair loss; human.
 XX OS Homo sapiens.
 XX PN WO200246465-A2.
 XX PD 13-JUN-2002.
 XX PF 10-DEC-2001; 2001WO-GB005458.
 XX PR 08-DEC-2000; 2000GB-00030076.
 XX PR 08-FEB-2001; 2001GB-00003156.
 XX PR 25-OCT-2001; 2001GB-00025666.
 XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;
 PI Rayner WN;
 XX WPI; 2002-627238/67.
 XX Identifying a gene involved in disease for treating hypoxia-regulated
 PT conditions, comprises comparing the transcriptome/proteome of two cell
 PT types under different conditions and identifying a differentially
 PT regulated gene.
 XX Claim 35; Page 397; 538pp; English.

The present invention relates to methods for identifying genes and
 proteins that are implicated in a specific disease or physiological
 condition. The method comprises comparing the transcriptome/proteome of a
 specialised cell type implicated in a disease or condition with that of a
 second specialised cell type, under two experimental conditions, and
 identifying a gene that is differentially regulated in the two
 specialised cell types under experimental conditions. ABV7873-ABV78116
 and ABP65061-ABP65257 were identified using the methods of the invention.
 The coding sequences and proteins are useful for treating a disease in a
 patient, for manufacture of a medicament for treating hypoxia-regulated
 conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
 CC biological response to hypoxia conditions, or hypoxic-associated
 CC pathology in a patient. The coding sequences and proteins are also useful
 CC for monitoring the therapeutic treatment of a disease or physiological
 CC condition, such as cancer, ischaemic conditions, reperfusion injury,
 CC retinopathy, neonatal stress, preclapmsia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss
 XX Sequence 193 AA;
 SQ

Query Match 100.0%; Score 1018; DB 5; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.9e-104;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIAGLLLATPAQAHLKPSQSSFSWNCDEGKDPVIRSLTLEPDPV 60
 Db 1 MQSLMQAPLLIAGLLLATPAQAHLKPSQSSFSWNCDEGKDPVIRSLTLEPDPV 60
 QY 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGCTFEHFCVDLMDLIP 120
 Db 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGCTFEHFCVDLMDLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPWLTGNVRIESVLSGSKR 180
 Db 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPWLTGNVRIESVLSGSKR 180
 QY 181 LGCIIKIAASLKG 193
 Db 181 LGCIIKIAASLKG 193

RESULT 9

ADN95858
 ID ADN95858 standard; protein; 193 AA.

XX AC ADN95858;
XX DT 01-JUL-2004 (first entry)
XX DE Human BEC/LEC-related protein sequence SeqID782.
XX KW growth; differentiation; blood endothelial cell; BEC;
XX KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
XX KW lymphatic growth agent; VEGF-D; angiogenic; cytosolic;
XX KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
XX KW inflammatory disease; cancer metastasis; lymphatic system; human.
XX OS Homo sapiens.
XX PN WO2003080640-A1.
XX PD 02-OCT-2003.
XX PF 07-MAR-2003; 2003WO-US006900.
XX PR 07-MAR-2002; 2002US-0363019P.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PA (LION) LICENTIA LTD.
XX PI Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
XX WPI; 2003-876899/81.
XX DR N-PSDB; ADN95859.
XX PS Example 1; SEQ ID NO 782; 176pp; English.
XX CC This invention relates to a method of differentially modulating the
CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
CC endothelial cells (LEC) comprising contacting endothelial cells with a
CC composition comprising an agent that differentially modulates blood or
CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
CC identifying a human subject with lymphoedema and with a mutation in at
CC least one allele of a gene encoding a LEC protein, where the mutation
CC correlates with lymphoedema in human subjects, and with the proviso that
CC the LEC protein is not VEGFR-3; and administering to the subject a
CC composition comprising a lymphatic growth agent selected from VEGF-C or
CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
CC the development of compounds with an angiogenic, cytosolic,
CC vasotropic or antiinflammatory activity or for gene therapy. The method
CC is useful in modulating the growth or differentiation of blood
CC endothelial cells or lymphatic endothelial cells, in treating hereditary
CC lymphoedema, in screening for an endothelial cell disorder or
CC predisposition to the disorder or in monitoring the efficacy or toxicity
CC of a drug on endothelial cells. The agent is useful in manufacturing a
CC medicament for the differential modulation of blood vessel endothelial
CC cell or lymphatic vessel endothelial cell growth or differentiation. The
CC lymphatic growth agent may also be used in manufacturing a medicament for
CC the treatment of hereditary lymphoedema resulting from a mutation in a
CC LEC gene or of other diseases involving the lymphatic vessels, such as
CC various inflammatory diseases and cancer metastasis via the lymphatic
CC system. The present sequence is that of a human LEC/BEC differentially
CC expressed protein which is related to the method of the invention. Note:
CC This sequence does not appear in the specification but was obtained by
CC the indexer using the source data given in table 14 of the specification.
XX SQ Sequence 193 AA;
Query Match 100.0%; Score 1018; DB 7; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.9e-104;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQSLMQAPLLIATGLLIATPAQHLKKPSQSSFSWNCDEGKDPVIRSLTLEPDPV 60
DB 1 MQSLMQAPLLIATGLLIATPAQHLKKPSQSSFSWNCDEGKDPVIRSLTLEPDPV 60
QY 61 PGNVTLVSGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDVLDMLIP 120

DB 61 PGNVTLVSGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDVLDMLIP 120
QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPWSLTTGNRYIESVLSGSKR 180
DB 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPWSLTTGNRYIESVLSGSKR 180
QY 181 LGCICKIAASLKGI 193
DB 181 LGCICKIAASLKGI 193
RESULT 10
ADN03620
ID ADN03620 standard; protein; 193 AA.
XX AC ADN03620;
XX DT 01-JUL-2004 (first entry)
XX DE Antipsoriatic protein sequence #7.
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX (GETH) GENENTECH INC.
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
XX Wu TD;
XX WPI; 2004-305105/28.
XX DR N-PSDB; ADN03619.
XX PT New PRO nucleic acid or polypeptide, useful for preparing a
XX pharmaceutical composition for diagnosing or treating psoriasis in a
XX mammal.
XX Claim 9; SEQ ID NO 14; 3069pp; English.
XX The invention relates to novel polynucleotide and polypeptides for
XX treating psoriasis or a sequence having at least 80% identity to the
XX above sequences. The nucleic acid is useful for preparing a composition
XX for diagnosing or treating psoriasis in a mammal. This sequence
XX corresponds to one of the polypeptides of the invention.
XX SQ Sequence 193 AA;
Query Match 100.0%; Score 1018; DB 8; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.9e-104;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQSLMQAPLLIATGLLIATPAQHLKKPSQSSFSWNCDEGKDPVIRSLTLEPDPV 60
DB 1 MQSLMQAPLLIATGLLIATPAQHLKKPSQSSFSWNCDEGKDPVIRSLTLEPDPV 60
QY 61 PGNVTLVSGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDVLDMLIP 120
DB 61 PGNVTLVSGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDVLDMLIP 120
QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPWSLTTGNRYIESVLSGSKR 180
DB 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPWSLTTGNRYIESVLSGSKR 180
QY 181 LGCICKIAASLKGI 193


```

Db      181 LGCIKIAASLKG1 193
|||||
RESULT 11
ABG00720
ID      ABG00720 standard; protein; 193 AA.
XX
XX
AC      ABG00720;
XX
XX      13-FEB-2002 (first entry)
XX
XX      Novel human diagnostic protein #711.
XX
XX      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW      food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX      Homo sapiens.
XX
XX      WO200175067-A2.
XX
XX      11-OCT-2001.
XX
XX      30-MAR-2001; 2001WO-US008631.
XX
XX      31-MAR-2000; 2000US-00540217.
XX
XX      23-AUG-2000; 2000US-00649167.
XX
XX      (HYSE-) HYSEQ INC.
XX
XX      Drmanac RT, Liu C, Tang YT;
XX
XX      WPI; 2001-639362/73.
XX
XX      N-PSDB; AAS64907.
XX
XX      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity.
XX
XX      Claim 20; SEQ ID NO 31079; 103pp; English.
XX
XX      The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC      sequences. (I) is useful as hybridisation probes, polymerase chain
CC      reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC      and in recombinant production of (II). The polynucleotides are also used
CC      in diagnostics as expressed sequence tags for identifying expressed
CC      genes. (I) is useful in gene therapy techniques to restore normal
CC      activity of (II) or to treat disease states involving (II). (II) is
CC      useful for generating antibodies against it, detecting or quantitating a
CC      polypeptide in tissue, as molecular weight markers and as a food
CC      supplement. (II) and its binding partners are useful in medical imaging
CC      of sites expressing (II). (I) and (II) are useful for treating disorders
CC      involving aberrant protein expression or biological activity. The
CC      polypeptide and polynucleotide sequences have applications in
CC      diagnostics, forensics, gene mapping, identification of mutations
CC      responsible for genetic disorders or other traits to assess biodiversity
CC      and to produce other types of data and products dependent on DNA and
CC      amino acid sequences. ABG00010-ABG030377 represent novel human diagnostic
CC      amino acid sequences of the invention. Note: The sequence data for this
CC      patent did not appear in the printed specification, but was obtained in
CC      electronic format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences
XX
XX      Sequence 193 AA;
SQ
Query Match      99.5%; Score 1013; DB 4; Length 193;
Best Local Similarity 99.5%; Pred. No. 6.7e-104;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 MQSLMQAPLLIAGLLATPAQAHLKXPSQSSFSWNCDEGKDPAVIRSLTLEPDP1WV 60
|||||
Db      1 MQSLMQAPLLIAGLLATPAQAHLKXPSQSSFSWNCDEGKDPAVIRSLTLEPDP1WV 60
|||||

```

QY 61 PGNVTLVSVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDVLDMLIP 120
 DB 61 PGNVTLVSVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDVLDMLIP 120
 QY 121 TGPCPEPLRTYGLPCHCPKEGTYSLPKSEFVVVDLELPSWLTGNYRIESVLSGSKR 180
 DB 121 TGPCPEPLRTYGLPCHCPKEGTYSLPKSEFVVVDLELPSWLTGNYRIESVLSGSKR 180
 QY 181 LGCIKIAASLKGI 193
 DB 181 LGCIKIAASLKGI 193
 RESULT 13
 ID AAB31899 standard; protein; 200 AA.
 AC AAB31899;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of a human protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 OS
 PN W0200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.
 XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 1; Page 160-161; 209pp; French.
 XX
 CC The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 200 AA;
 Query Match 99.1%; Score 1009; DB 4; Length 200;
 Best Local Similarity 98.4%; Pred. No. 2e-103;
 Matches 190; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MQSIMQAPLLIALGLLLATPAQAHLLKPSQLSFSWDCDEGKDPVIRSLTLEPDPPIV 60
 DB 8 MQSIMQAPLLIALGLLLAQAHLKPSQLSFSWDCDEGKDPVIRSLTLEPDPPIV 67
 QY 61 PGNVTLVSVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDVLDMLIP 120
 DB 68 PGNVTLVSVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDVLDMLIP 127
 QY 121 TGPCPEPLRTYGLPCHCPKEGTYSLPKSEFVVVDLELPSWLTGNYRIESVLSGSKR 180
 DB 128 TGPCPEPLRTYGLPCHCPKEGTYSLPKSEFVVVDLELPSWLTGNYRIESVLSGSKR 187
 QY 181 LGCIKIAASLKGI 193
 DB 188 LGCIKIAASLKGI 200
 RESULT 14
 ID AAB31897 standard; protein; 193 AA.
 AC AAB31897;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of a mutant ganglioside GM2 activator protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 OS
 PN W0200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.
 XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 25; Page 159-160; 209pp; French.
 XX
 CC The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 98.7%; Score 1005; DB 4; Length 193;
Best Local Similarity 99.0%; Pred. No. 5.2e-103;
Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIALGLLLATPAQAHLKKPSQSFSDNCDEGKDPVIRSLTLEPDPPIV 60
DB 1 MQSLMQAPLLIALGLLLATPAQAHLKKPSQSFSDNCDEGKDPVIRSLTLEPDPPIV 60
QY 61 PGNVTLVVVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
DB 61 PGNVTLVVVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVDLELPSWLTTCNRYRIESVLSGSKR 180
DB 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVDLELPSWLTTCNRYRIESVLSGSKR 180
QY 181 LGCIKIAASLKG I 193
DB 181 LGCIKIAASLKG I 193

RESULT 15
AAB31900
ID AAB31900 standard; protein; 189 AA.
XX
AC AAB31900;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
FN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INNR) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX
DR WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
FS Claim 1; Page 161-162; 209pp; French.
XX

Search completed: November 17, 2004, 06:12:05
Job time : 34.4237 secs

CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 189 AA;

Query Match 98.2%; Score 1000; DB 4; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.8e-102;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 MQAPLLIALGLLLATPAQAHLKKPSQSFSDNCDEGKDPVIRSLTLEPDPPIVVPGNV 64
DB 1 MQAPLLIALGLLLATPAQAHLKKPSQSFSDNCDEGKDPVIRSLTLEPDPPIVVPGNV 60
QY 65 TLSVVVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLPTGEP 124
DB 61 TLSVVVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLPTGEP 120
QY 125 CPEPLRTYGLPCHCPKFGTYSLPKSEFVVDLELPSWLTTCNRYRIESVLSGSKRLGCI 184
DB 121 CPEPLRTYGLPCHCPKFGTYSLPKSEFVVDLELPSWLTTCNRYRIESVLSGSKRLGCI 180
QY 185 KIAASLKG I 193
DB 181 KIAASLKG I 189

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 05:44:53 ; Search time 7.65624 Seconds
(without alignments)
1671.758 Million cell updates/sec

Title: US-10-030-937-8
Perfect score: 1018
Sequence: 1 MQSLMQAPLLIALGILLATP.....LSSSGKRLGCIXIASLKGI 193

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/iaa/5A COMB.pap:*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pap:*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pap:*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pap:*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pap:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1018	100.0	193	4	US-09-183-841-1
2	870	85.5	178	4	US-09-183-841-2
3	86.5	8.5	410	3	US-08-630-172-17
4	86.5	8.5	410	3	US-09-375-419-17
5	86	8.4	143	4	US-09-860-793-3
6	85.5	8.4	310	4	US-09-976-594-807
7	83	8.2	746	4	US-09-370-838-185
8	83	8.2	746	4	US-09-854-133-185
9	82	8.1	572	6	5256770-7
10	81	8.0	446	1	US-08-307-444A-5
11	81	8.0	446	1	US-08-587-389-5
12	81	8.0	456	1	US-08-307-444A-3
13	81	8.0	456	1	US-08-307-444A-4
14	81	8.0	456	1	US-08-587-389-3
15	81	8.0	456	1	US-08-587-389-4
16	81	8.0	475	1	US-08-307-444A-1
17	81	8.0	475	1	US-08-307-444A-2
18	81	8.0	475	1	US-08-587-389-1
19	81	8.0	475	1	US-08-587-389-2
20	81	8.0	476	1	US-08-014-723-1
21	81	8.0	476	1	US-08-014-723-2
22	81	8.0	476	1	US-08-014-723-18
23	81	8.0	476	1	US-08-110-011A-1
24	81	8.0	476	1	US-08-110-011A-2
25	81	8.0	476	1	US-08-110-011A-18
26	81	8.0	494	1	US-08-014-723-14
27	81	8.0	494	1	US-08-014-723-16

28	81	8.0	494	1	US-08-110-011A-14	Sequence 14, Appl
29	81	8.0	494	1	US-08-110-011A-16	Sequence 16, Appl
30	81	8.0	497	1	US-08-312-870-3	Sequence 3, Appl
31	81	8.0	497	4	US-09-331-793-4	Sequence 4, Appl
32	81	8.0	498	2	US-08-733-564-2	Sequence 2, Appl
33	81	8.0	575	1	US-08-261-206A-59	Sequence 59, Appl
34	81	8.0	575	1	US-08-312-870-1	Sequence 1, Appl
35	81	8.0	575	1	US-08-170-290A-54	Sequence 54, Appl
36	81	8.0	575	4	US-09-880-484D-2	Sequence 2, Appl
37	81	8.0	575	4	US-10-438-648-2	Sequence 2, Appl
38	81	8.0	575	6	5466668-6	Patent No. 5466668
39	80.5	7.9	302	4	US-09-248-796A-14539	Sequence 14539, A
40	77.5	7.6	126	4	US-09-860-793-1	Sequence 1, Appl
41	77.5	7.6	768	3	US-09-302-812-8	Sequence 8, Appl
42	77.5	7.6	768	3	US-09-511-477-8	Sequence 8, Appl
43	77.5	7.6	768	3	US-09-511-507-8	Sequence 8, Appl
44	75.5	7.4	269	3	US-09-082-649B-82	Sequence 82, Appl
45	75.5	7.4	269	3	US-09-430-503-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-183-841-1
; Sequence 1, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (33)..(55)
; FEATURE:
; OTHER INFORMATION: residues 56-63 are included in a further precursor
; OTHER INFORMATION: form of the protein
US-09-183-841-1

Query Match		100.0%	Score 1018:	DB 4:	Length 193;
Best Local Similarity		100.0%	Pred. No. 7.1e-113;		
Matches 193;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MQSLMQAPLLIALGILLATPAAHLKPKS	QSLSSFSWNCDEGKDPVIRSLTLEPDP	IVV	60
Db	1	MQSLMQAPLLIALGILLATPAAHLKPKS	QSLSSFSWNCDEGKDPVIRSLTLEPDP	IVV	60
Qy	61	PGNVTLSVVGSTVPLSSPLKVDLVLEK	VAGLWIKIPCTDIYIGSCTFEHFC	CDVLDMLIP	120
Db	61	PGNVTLSVVGSTVPLSSPLKVDLVLEK	VAGLWIKIPCTDIYIGSCTFEHFC	CDVLDMLIP	120
Qy	121	TGPEPCPEPLRTYGPCHCPKKEGYSL	PKSFVVDLPLPSWLTGNNVRIESV	SSSGKR	180
Db	121	TGPEPCPEPLRTYGPCHCPKKEGYSL	PKSFVVDLPLPSWLTGNNVRIESV	SSSGKR	180
Qy	181	LGCIXIASLKGI	193		
Db	181	LGCIXIASLKGI	193		

RESULT 2
US-09-183-841-2
; Sequence 2, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:

; APPLICANT: Hospital for Sick Children
 ; TITLE OF INVENTION: A No. 642368061 Inhibitor of Platelet Activating Factor
 ; FILE REFERENCE: van0010
 ; CURRENT APPLICATION NUMBER: US/09/183,841
 ; CURRENT FILING DATE: 1998-10-30
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 178
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: His tag at residues 1 to 17
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: amino acid
 ; OTHER INFORMATION: sequence of Gm2 protein using His6 tag
 US-09-183-841-2

Query Match 85.5%; Score 870; DB 4; Length 178;
 Best Local Similarity 100.0%; Pred. No. 2.6e-95; Indels 0; Gaps 0;
 Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 32 SFSWNCDEGKDPVIRSLTLEPDPPIVPGNVTLVWGSTVPLSPLKVDLVLEKEVA 91
 Db 17 SFSWNCDEGKDPVIRSLTLEPDPPIVPGNVTLVWGSTVPLSPLKVDLVLEKEVA 76
 QY 92 GLWIKIPCTDIYIGSCFTEHPCDVLMLPTGEPCEPRLRYGLPCHGPFKEGTYSLPKSE 151
 Db 77 GLWIKIPCTDIYIGSCFTEHPCDVLMLPTGEPCEPRLRYGLPCHGPFKEGTYSLPKSE 136
 QY 152 FVVPDLPLSWLTGNRIEIVSSSGKRLGCIKIASLKG1 193
 Db 137 FVVPDLPLSWLTGNRIEIVSSSGKRLGCIKIASLKG1 178

RESULT 3

US-08-630-172-17
 ; Sequence 17, Application US/08630172
 ; Patent No. 6060054
 ; GENERAL INFORMATION:
 ; APPLICANT: Staerz, Uwe
 ; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
 ; TITLE OF INVENTION: LYMPHOCTE VETO
 ; NUMBER OF SEQUENCES: 41
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sheridan Ross & McIntosh
 ; STREET: 1700 Lincoln Street, 35th Floor
 ; CITY: Denver
 ; STATE: Colorado
 ; COUNTRY: U.S.
 ; ZIP: 80203
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/630,172
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Connell, Gary J.
 ; REGISTRATION NUMBER: 32,020
 ; REFERENCE/DOCKET NUMBER: 2879-36
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (303) 863-9700
 ; TELEFAX: (303) 863-0223
 ; INFORMATION FOR SEQ ID NO: 17:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 410 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

US-08-630-172-17

Query Match 8.5%; Score 86.5; DB 3; Length 410;
 Best Local Similarity 25.6%; Pred. No. 0.17;
 Matches 42; Conservative 14; Mismatches 43; Indels 65; Gaps 9;
 QY 26 KKPESQLSSFS-----WDCNDEGKDPVIRSLTLEPDPPIV-----PGNVTLVWGST- 72
 Db 46 KGPSKLNDRADRSRLW---DQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGTA 102
 QY 73 -----SVPLSSP-----LKVDLVLEKEVAGLWIKIP 98
 Db 103 NSDTHLLQGQSLTLTLESPPGSPVQCRSPRGKNIQGGKTLVS-QLELQDAGTWI--- 158
 QY 99 CTDYIGSCFTEHPCDVLMLPTG---EPCPEPLRYGLPCHCP 139
 Db 159 CTVLQNGKVKVEFKIDIVLAEPRGPTIKPCP-----PCKGP 194

RESULT 4

US-09-375-419-17
 ; Sequence 17, Application US/09375419
 ; Patent No. 6264950
 ; GENERAL INFORMATION:
 ; APPLICANT: Staerz, Uwe
 ; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
 ; TITLE OF INVENTION: LYMPHOCTE VETO
 ; NUMBER OF SEQUENCES: 41
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sheridan Ross & McIntosh
 ; STREET: 1700 Lincoln Street, 35th Floor
 ; CITY: Denver
 ; STATE: Colorado
 ; COUNTRY: U.S.
 ; ZIP: 80203
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/375,419
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/630,172
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Connell, Gary J.
 ; REGISTRATION NUMBER: 32,020
 ; REFERENCE/DOCKET NUMBER: 2879-36
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (303) 863-9700
 ; TELEFAX: (303) 863-0223
 ; INFORMATION FOR SEQ ID NO: 17:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 410 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

Query Match 8.5%; Score 86.5; DB 3; Length 410;
 Best Local Similarity 25.6%; Pred. No. 0.17;
 Matches 42; Conservative 14; Mismatches 43; Indels 65; Gaps 9;
 QY 26 KKPESQLSSFS-----WDCNDEGKDPVIRSLTLEPDPPIV-----PGNVTLVWGST- 72
 Db 46 KGPSKLNDRADRSRLW---DQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGTA 102
 QY 73 -----SVPLSSP-----LKVDLVLEKEVAGLWIKIP 98
 Db 103 NSDTHLLQGQSLTLTLESPPGSPVQCRSPRGKNIQGGKTLVS-QLELQDAGTWI--- 158

```

QY 99 CTDYIGCTFEHFCVDLMDLPTG---EPDCEPLRTYGLPCHCP 139
Db 159 CTVLQNKQKVEFKIDIVVLAEPRGFTIKPCP-----PCKCP 194

RESULT 5
US-09-860-793-3
; Sequence 3, Application US/09860793
; Patent No. 6559121
; GENERAL INFORMATION:
; APPLICANT: Pruett, John H
; APPLICANT: Temeyer, Kevin B
; APPLICANT: Kunz, Sidney E
; APPLICANT: Fisher, William F
; TITLE OF INVENTION: Vaccines for the Protection of Cattle from Psoroptic
; FILE OF INVENTION: Scabies
; FILE REFERENCE: Docket 0047.96 - John H. Pruett et al.
; CURRENT APPLICATION NUMBER: US/09/860,793
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/366,603
; PRIOR FILING DATE: 1999-08-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Psoroptes ovis
US-09-860-793-3

Query Match      8.4%; Score 86; DB 4; Length 143;
Best Local Similarity 22.4%; Pred. No. 0.04;
Matches 44; Conservative 30; Mismatches 50; Indels 72; Gaps 11;

QY 10 LIALGLLILATPAQAHLKKPSQSFSDNCDGKDPVIRSLTLE---PDPVIVPGNVTL 66
Db 5 LVLAITLAVVSAGKVK-----PQDCGKG-----VESLEVGCGSDYCVIHKGKGL 51

QY 67 SV-VGTSVPLSPKVDLVLEKEVAGLWIKIPCTDYIGCTFEHFCVDLMDLPTGEP 125
Db 52 DLAISVTSNQDSANLKLDIV--ADINGVQIEVPGVDHGD----- 88

QY 126 PEPLRTYGLPCH---CPKEG-----TYSLPKSEFVVPDLEPLSMITGNYRIESVLS 175
Db 89 -----CHYVKCPKKGQHFVDVYTSIPA---ILP-----TTKAKIIAKIIG 127

QY 176 SSGKRLGCIKIAASLK 191
Db 128 DKGLG-GCIVNGEIQ 142

RESULT 6
US-09-976-594-807
; Sequence 807, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 807
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 2770104C1D

```

US-09-976-594-807

```

Query Match      8.4%; Score 85.5; DB 4; Length 310;
Best Local Similarity 24.7%; Pred. No. 0.14;
Matches 55; Conservative 20; Mismatches 67; Indels 81; Gaps 10;

QY 9 LLIATGLLILATPAQAHL---LKKPSQSFSDNCDGKDPVIRSLTLEPPIVPGN-V 64
Db 51 VLICGLSLGRTHVQTGTPKP-----TLWAEPSDVTIQSSPV 89

QY 65 TLSVVGSTSVPLSPKVDLVLEKEVAGLWIKIPCTD-----YIGSCTFEH----- 110
Db 90 TLSQGSLEAQ-----EYRLYREKKSAS-WLIRPELVKNGQPHISITWEHTGRVGCQ 143

QY 111 -----FCDVLDMLI-----PTGCPCEPLRTYV-----LPCHCFKEGTYSLPKSE 151
Db 144 YYSRARSSELSDPLVAGDRSYQNTQSPGPGVVTGKNVTLCCQSRGQPHTELLTKEG 203

QY 152 FVVPDLEPLS-----WLTGNGYRIESVLS 177
Db 204 AGHPHLRSEHQAOQNAEFMRMGVTSAHVGTIRCYSLSN 246

```

RESULT 7

```

US-09-370-838-185
; Sequence 185, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 185
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-185

```

```

Query Match      8.2%; Score 83; DB 4; Length 746;
Best Local Similarity 25.2%; Pred. No. 1;
Matches 54; Conservative 15; Mismatches 79; Indels 66; Gaps 11;

```

```

QY 21 AQAH-LKKPSQSFSDNCDG---KDPVIRSLTLEPPIVVPNGVTLVVGSTSVPL 76
Db 257 SQKHSLSQSQISPKSWGATASLIPNDQLLPKLNTEPKDVPKPVH---QPVGSG----- 308

QY 77 SSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMD-----LIPTGEP----- 124
Db 309 SSTLPKDPVLRKEK-----LQDLMTQIQGTCNFMQES-SVLDPDKFSSAIPTSQPPSATPGS 363

QY 125 -----CPEPLRTYGLPCHCP-----PKEGTYS-----LPKSEFVVPD 156
Db 364 PVASKEQNLSQSDFLQEPLOQVFNVAFLPPRKEQEKESFSPGYNQSFSTASTQTTPQ 423

QY 157 LELP-----SWLTTGNYRIESVLSSSGKRL 181
Db 424 CQLPSIHVEQTVHVSQETAANYHPDGTIQVNSGL 457

```

RESULT 8

```

US-09-854-133-185
; Sequence 185, Application US/09854133
; Patent No. 6759508
; GENERAL INFORMATION:

```

```

; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 185
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-854-133-185

```

```

Query Match      8.2%; Score 83; DB 4; Length 746;
Best Local Similarity 25.2%; Pred. No. 1;
Matches 54; Conservative 15; Mismatches 79; Indels 66; Gaps 11;

QY 21 AQAH-LKPSQSSFSWDCNDEG---KDPVIRSLTLEPDPVIVPGNVTLVVVGSSTVPL 76
Db 257 SQKSLTSQSIQSPKSGVATASLIENDQLLPKLNTEPKDVPKH---QPVGS----- 308
QY 77 SPSLKVDLVLEKAVAGLWIKIPCTDYIGSCTFEHFCVDLDM-----LIPTGEP----- 124
Db 309 SSTLPKDPVLRKEK---LQDLMTQIQGTGNFQW---SVLDFDKPSAIPTSQPPSATPGS 363
QY 125 -----CPEPLRTVGLPCHCP-----FKSGTYS-----LPSKSEFVVD 156
Db 364 PVASKEQNLSSQDFLOEPQVFNNAVLPPRKEQIKSPSPGYNQSFSTASTQTTPQ 423
QY 157 LELP-----SWLTGNRYIESVLSGSKRL 181
Db 424 CQLPSIHVQTVHSQETAANYHPDGTIQVNGSL 457

```

```

RESULT 9
5256770-7
; Patent No. 5256770
; APPLICANT: GLASER, CHARLES B.; MORSE, MICHAEL J.; LIGHT,
; DAVID R.
; TITLE OF INVENTION: OXIDATION RESISTANT THROMBOMODULIN ANALOGS
; NUMBER OF SEQUENCES: 48
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/506,325
; FILING DATE: 09-APR-1990
; SEQ ID NO: 7;
; LENGTH: 572
5256770-7

```

```

Query Match      8.1%; Score 82; DB 6; Length 572;
Best Local Similarity 23.6%; Pred. No. 0.93;
Matches 49; Conservative 23; Mismatches 74; Indels 62; Gaps 12;

QY 8 PLLIALGLLLAT-----PAAHLKPKSQSSFSWDCNDEGKDPVIRSLTLEPDPV 59
Db 134 PLCVAVSAAEATVPSEPIWEEQCEVKADGFLCEFH-----PATCRPLAVEPGAA 185
QY 60 VFGNVTLV-----VGSTSVPLSSPLKVDLVL-----EKEVAGLWIKI 97
Db 186 AAVSITYGTPFAARGADFOALPVGSSAA--VAPLGLQMLCTAGNVGHWAREAPGAW--- 240
QY 98 PCTDYIGSCTFEHFCVDLMDLIPGCPPEP-----LRTYGLPCHCPFKGTYSLPKSEFV 153
Db 241 DCSVENGGC--EHACNA----IPGAPRCQCPAGAAQADGRSCTASATQSCNDLCE-HFC 293
QY 154 VPDLELPSWLTGNRYIESVLSGSKRL 181
Db 294 VFNPDQP-----GSY---SCMCEGTGYRL 313

```

```

RESULT 10
US-08-307-444A-5
; Sequence 5, Application US/08307444A
; Patent No. 5516659
; GENERAL INFORMATION:
; APPLICANT: NIL, ATSUSHI
; APPLICANT: MORISHITA, HIDEAKI
; APPLICANT: UEMURA, AKIO
; APPLICANT: MOCHIDA, EI
; TITLE OF INVENTION: ANTICOAGULANT POLYPEPTIDES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,444A
; FILING DATE: 19-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/835,436
; FILING DATE: 26-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: OLIFF, JAMES A.
; REGISTRATION NUMBER: 27,075
; REFERENCE/DOCKET NUMBER: JAO 27706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6400
; TELEFAX: (703) 836-2787
; TELEX: 90-1799 PTO ALEX
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-307-444A-5

```

```

Query Match      8.0%; Score 81; DB 1; Length 446;
Best Local Similarity 23.3%; Pred. No. 0.85;
Matches 49; Conservative 23; Mismatches 74; Indels 64; Gaps 12;

QY 8 PLLIALGLLLAT-----PAAHLKPKSQSSFSWDCNDEGKDPVIRSLTLEPDPV 59
Db 117 PLCVAVSAAEATVPSEPIWEEQCEVKADGFLCEFH-----PATCRPLAVEPGAA 168
QY 60 VFGNVTLV-----VGSTSVPLSSPLKVDLVL-----EKEVAGLWIKI 95
Db 169 AAVSITYGTPFAARGADFOALPVGSSAA--VAPLGLQMLCTAGNVGHWAREAPGAW--- 225
QY 96 KIPCTDYIGSCTFEHFCVDLMDLIPGCPPEP-----LRTYGLPCHCPFKGTYSLPKSE 151
Db 226 --DCSVENGGC--EHACNA----IPGAPRCQCPAGAAQADGRSCTASATQSCNDLCE-H 276
QY 152 FVVPDLELPSWLTGNRYIESVLSGSKRL 181
Db 277 FCVPNDQP-----GSY---SCMCEGTGYRL 298

RESULT 11
US-08-587-389-5

```



```

; Sequence 5, Application US/08587389
; Patent No. 5695964
; GENERAL INFORMATION:
; APPLICANT: NII, ATSUSHI
; APPLICANT: MORISHITA, HIDEAKI
; APPLICANT: UEMURA, AKIO
; APPLICANT: MOCHIDA, EI
; APPLICANT: UEMURA, AKIO
; TITLE OF INVENTION: TRUNCATED THROMBOMODULIN, RECOMBINANT
; TITLE OF INVENTION: PRODUCTION THEREOF, AND THERAPEUTIC AGENT (AS AMENDED)
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/587,389
; FILING DATE: 17-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/307,444
; FILING DATE: 19-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: OLIFF, JAMES A.
; REGISTRATION NUMBER: 27,075
; REFERENCE/DOCKET NUMBER: JAO 27706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6400
; TELEFAX: (703) 836-2787
; TELEX: 90-1799 PTO ALEX
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-587-389-5

Query Match      8.0%; Score 81; DB 1; Length 446;
Best Local Similarity 23.3%; Pred. No. 0.85;
Matches 49; Conservative 23; Mismatches 74; Indels 64; Gaps 12;

QY      8 PLLIALGLLLAT-----PAQHLKKPSQLSSFSWDCDEGKDPVAVIRSLTLEPDPV 59
Db      117 PLCVAVSAAEATVPSEPIWEQCEVKADGFLCFEHP-----PATCRPLAVEPGAA 168

QY      60 VPGNVTLVS-----VGSTSVPLSPLKVDLVL-----EKEVAGLWI 95
Db      169 AAVSITYGTPFAARGADFOALPVGSSAA--VAPLGLQLMCTAPPAGVQGHWAREAPGAW- 225

QY      96 KIPCTDYIGSCFTFFHCDVLDMLIPTGEPCEP-----LRTYGLPCHCFKGTYSLPKSE 151
Db      226 --DCSVENGCC--EHACNA----IPGAPRCQCPAGAAQADGRSCTASATSCNDLCB-H 276

QY      152 FVVPDLPLPSWLTGNYRIESVLSSSGKRL 181
Db      277 FCVFPNDQP-----GSY---SCMCETGYRL 298

RESULT 12
US-08-307-444A-3
; Sequence 3, Application US/08307444A
; Patent No. 5516659

```

```

; GENERAL INFORMATION:
; APPLICANT: NII, ATSUSHI
; APPLICANT: MORISHITA, HIDEAKI
; APPLICANT: UEMURA, AKIO
; APPLICANT: MOCHIDA, EI
; TITLE OF INVENTION: ANTICOAGULANT POLYPEPTIDES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,444A
; FILING DATE: 19-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/835,436
; FILING DATE: 26-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: OLIFF, JAMES A.
; REGISTRATION NUMBER: 27,075
; REFERENCE/DOCKET NUMBER: JAO 27706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6400
; TELEFAX: (703) 836-2787
; TELEX: 90-1799 PTO ALEX
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-307-444A-3

Query Match      8.0%; Score 81; DB 1; Length 456;
Best Local Similarity 23.3%; Pred. No. 0.87;
Matches 49; Conservative 23; Mismatches 74; Indels 64; Gaps 12;

QY      8 PLLIALGLLLAT-----PAQHLKKPSQLSSFSWDCDEGKDPVAVIRSLTLEPDPV 59
Db      117 PLCVAVSAAEATVPSEPIWEQCEVKADGFLCFEHP-----PATCRPLAVEPGAA 168

QY      60 VPGNVTLVS-----VGSTSVPLSPLKVDLVL-----EKEVAGLWI 95
Db      169 AAVSITYGTPFAARGADFOALPVGSSAA--VAPLGLQLMCTAPPAGVQGHWAREAPGAW- 225

QY      96 KIPCTDYIGSCFTFFHCDVLDMLIPTGEPCEP-----LRTYGLPCHCFKGTYSLPKSE 151
Db      226 --DCSVENGCC--EHACNA----IPGAPRCQCPAGAAQADGRSCTASATSCNDLCB-H 276

QY      152 FVVPDLPLPSWLTGNYRIESVLSSSGKRL 181
Db      277 FCVFPNDQP-----GSY---SCMCETGYRL 298

RESULT 13
US-08-307-444A-4
; Sequence 4, Application US/08307444A
; Patent No. 5516659
; GENERAL INFORMATION:
; APPLICANT: NII, ATSUSHI
; APPLICANT: MORISHITA, HIDEAKI

```

RESULT 14
US-08-587-389-3
; Sequence 3, Application US/08587389
; Patent No. 5635964
; GENERAL INFORMATION:
; APPLICANT: NII, ATSUSHI
; APPLICANT: MORISHITA, HIDEAKI
; APPLICANT: UEMURA, AKIO
; APPLICANT: MOCHIDA, EI
; TITLE OF INVENTION: TRUNCATED THROMBOMODULIN, RECOMBINANT

RESULT 15
US-08-587-389-4
; Sequence 4, Application US/08587389
; Patent No. 5695964
; GENERAL INFORMATION:
; APPLICANT: NII, ATSUSHI
; APPLICANT: MORISHITA, HIDEAKI
; APPLICANT: UEMURA, AKIO
; APPLICANT: MOCHIDA, EI
; TITLE OF INVENTION: TRUNCATED THROMBOMODULIN, RECOMBINANT
; TITLE OF INVENTION: PRODUCTION THEREOF, AND THERAPEUTIC AGENT (AS AMENDED)
; NUMBER OF SEQUENCES: 27

;;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/587,389
; FILING DATE: 17-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/307,444
; FILING DATE: 19-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: OLIFF, JAMES A.
; REGISTRATION NUMBER: 27,075
; REFERENCE/DOCKET NUMBER: JAO 27706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6400
; TELEFAX: (703) 836-2787
; TELEX: 90-1799 PTO ALEX
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-587-389-4

Query Match 8.0%; Score 81; DB 1; Length 456;
Best Local Similarity 23.3%; Pred. No. 0.87;
Matches 49; Conservative 23; Mismatches 74; Indels 64; Gaps 12;

Qy 8 PLLTALGILLAT-----PAAHLKKPSQLSSFSWNCDEGKDPVIRSLTLEPDPV 59
Db 117 PLCVAVSAEATVPSEPIWEQQCEVKADGFLCEHF-----PATCRPLAVEPGAA 168

Qy 60 VPGNVTLSV-----VGSTSVPLSSPLKVDLVL-----EKEVAGLWI 95
Db 169 AAVSITYGTPFAARGADPQALFVGSSAA--VAPLGQLMCTAPPQAVQGHWAREAPGAW- 225

Qy 96 KIPCTDYIGSCTFEHCVDVLDMLIPTGBPCPEP----LRTYGLPCHCFKGTYSLPKSE 151
Db 226 --DCSVENGCC--EHACNA----IPGAPRCQCPAGAALQADGRSCTAGATSCNDLCE-H 276

Qy 152 FVVPDLPLPSMLTTGNRYRIESVLSGSKRL 181
Db 277 FCVPNPQDP-----GSY----SCMCETGYRL 298

Search completed: November 17, 2004, 06:43:23
Job time : 8.70624 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 05:48:19 ; Search time 25.0941 Seconds
(without alignments)
2723.613 Million cell updates/sec

Title: US-10-030-937-8

Perfect score: 1018

Sequence: 1 MQSLMQAPLLIALGILLATP.....LSSGKRLGCIKIAASLKG1 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pcp.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1018	100.0	193	14	US-10-170-385-389
2	411	40.4	191	15	US-10-264-049-2611
3	353	34.7	61	9	US-09-864-761-34809
4	162.5	16.0	126	10	US-09-764-891-4977
5	97.5	9.6	273	17	US-10-425-115-297138
6	96.5	9.5	641	14	US-10-369-493-20746
7	88.5	8.7	196	17	US-10-425-115-251307
8	87.5	8.6	152	16	US-10-437-963-149516
9	86	8.4	143	9	US-09-860-793-3
10	86	8.4	651	14	US-10-369-493-20109
11	85.5	8.4	574	17	US-10-725-013-2
12	85	8.3	575	15	US-10-094-886-196
13	85	8.3	5546	15	US-10-296-734-1210

14 84.5 8.3 247 17 US-10-425-115-279555
15 84.5 8.3 342 17 US-10-425-115-220322
16 84.5 8.3 342 17 US-10-425-115-220323
17 84.5 8.3 352 15 US-10-425-114-63175
18 84.5 8.3 371 15 US-10-425-114-62424
19 84.5 8.3 372 15 US-10-425-114-63454
20 84.5 8.3 372 15 US-10-425-114-72059
21 84.5 8.3 375 15 US-10-425-114-58493
22 84 8.3 611 14 US-10-369-493-7417
23 84 8.3 633 14 US-10-369-493-4658
24 83 8.2 688 15 US-10-282-122A-47772
25 83 8.2 743 14 US-10-309-422-34
26 83 8.2 746 9 US-09-738-973-185
27 83 8.2 746 9 US-09-854-133-185
28 83 8.2 746 14 US-10-144-649A-185
29 83 8.2 877 14 US-10-309-422-32
30 83 8.2 908 14 US-10-309-422-10
31 83 8.2 909 14 US-10-309-422-22
32 83 8.2 1042 14 US-10-309-422-8
33 83 8.2 1043 14 US-10-309-422-20
34 82 8.1 157 15 US-10-424-599-195822
35 82 8.1 166 15 US-10-424-599-210949
36 82 8.1 474 14 US-10-369-493-19074
37 81.5 8.0 152 16 US-10-767-701-46035
38 81.5 8.0 651 14 US-10-369-493-12040
39 81 8.0 255 16 US-10-437-963-187582
40 81 8.0 497 14 US-10-298-796-4
41 81 8.0 575 9 US-09-938-405-2
42 81 8.0 575 10 US-09-880-464-2
43 81 8.0 575 14 US-10-150-440-3
44 81 8.0 575 14 US-10-438-648-2
45 81 8.0 575 15 US-10-410-195-2

ALIGNMENTS

RESULT 1

US-10-170-385-389
; Sequence 389, Application US/10170385
; Publication No. US20030203372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 389
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-170-385-389

Query Match 100.0%; Score 1018; DB 14; Length 193;
Best Local Similarity 100.0%; Pred. No. 1e-97;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIALGILLATPAAHLKPKPSQSWNCDEGKDPVIRSLTLEPDP1V 60

```

Db      1  MQSLMQAPLIAUGLLATPAQAHKKPSQSSFSMDNCDEGKDPVAVIRSLTLEPPIV 60
QY      61  PGNVTLVSVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDVLDMLIP 120
Db      61  PGNVTLVSVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDVLDMLIP 120
QY      121  TGPFCPEPLRTYGLPCHCPKPGTYSILPKSEFVVPDLELPSWLTGTYNRYIESVLSSGKR 180
Db      121  TGPFCPEPLRTYGLPCHCPKPGTYSILPKSEFVVPDLELPSWLTGTYNRYIESVLSSGKR 180
QY      181  LGCICKIAASLKGI 193
Db      181  LGCICKIAASLKGI 193

RESULT 2
US-10-264-049-2611
; Sequence 2611, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133p1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: Patentin ver. 3.1
; SEQ ID NO 2611
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (141)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (142)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (184)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (187)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2611

Query Match      40.4%; Score 411; DB 15; Length 191;
Best Local Similarity 48.4%; Pred. No. 2e-34;
Matches 78; Conservative 24; Mismatches 51; Indels 8; Gaps 2;

QY      13  LGLLLA-----TPAQLKKPSQSSFSMDNCDEGKDPVAVIRSLTLEPPIVPGNVT 65
Db      24  LGLLLAGPAAHAVPAHAPVPPQVTSFFWENCHERKDPVLKSMLEPDPPIAYPGNVT 83
QY      66  LSVVGSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCCTFEHFCVDVLDMLIPTGEP 125
Db      84  ISAEQLQVRVPLSSPQKVELIEKVFVWFKVPCMSHV-RCIFDICIQLDFLIPPGQXX 142
QY      126  PEPLRTYGLPCHCPKPGTYSILPKSEFVVPDLELPSWLTG 166
Db      143  PEPLRTYGLPCTVSSRHLLNAQRILKPCFNTDLPGLITSG 183

RESULT 3
US-09-864-761-34809

```

```

; Sequence 34809, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34809
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011342.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99
; OTHER INFORMATION: EST HUMAN HIT: BE18286.1, EVALUE 9.00e-34
; OTHER INFORMATION: SWISSPROT HIT: P17900, EVALUE 1.00e-34
US-09-864-761-34809

Query Match      34.7%; Score 353; DB 9; Length 61;
Best Local Similarity 100.0%; Pred. No. 5.1e-29;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 82 VDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLPTGECPCPEPLRTYGLPCHCPK 141
Db 1 VDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLPTGECPCPEPLRTYGLPCHCPK 60

QY 142 E 142

Db 61 E 61

RESULT 4

US-09-764-891-4977
; Sequence 4977, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4977
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (119)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (122)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (123)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4977

Query Match 16.0%; Score 162.5; DB 10; Length 126;
Best Local Similarity 54.2%; Pred. No. 9.8e-09;
Matches 32; Conservative 7; Mismatches 13; Indels 7; Gaps 1;

QY 13 LGLLIA-----TPAQHLKPSQLSFSWDCDEGKDPVIRSLTEPDPPIVPGNV 64
Db 24 LGLLIAGPAHAHVPAHAPVNPQVVISFFMENCHERKDPVLLKMTLEPDPPIAYPGNV 82

RESULT 5

US-10-425-115-297138
; Sequence 297138, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 297138
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_34069C.1.pap
US-10-425-115-297138

Query Match 9.6%; Score 97.5; DB 17; Length 273;
Best Local Similarity 26.7%; Pred. No. 0.16;

Matches 59; Conservative 24; Mismatches 73; Indels 65; Gaps 14;
QY 3 SLMOAPL---LVALGILLATPAQHLK-----PSQLSFSWDCDEGKDPVIRSLTEPDPPIVPGNV 141
Db 70 ALLPAALPASLPPAASILYPLQAVLPRCGISLCPAELAPA--PSCSSRPFPAPLLAWSS 127
QY 51 LTLEPDPPIVPGNVTLVSVVSTVPLSSPLKY-----DLVLEKEVAGLW----- 94
Db 128 SFLVPVPCVFP--VTSMAAPSL--FLSSPLRVPLVSSLRARSFLCVPAATSLWCLFLPA 184
QY 95 -IKIPCTDYIGSCTFEHFCVDLMDLPTGECPCPEPLRTYGLPCH-----CPFKEGTY 145
Db 185 RAKFFCSIALGPC-----STVPCCFSARV-KFPCRVCLGRKPKVCPRR--AC 227
QY 146 SLPKSEFVVVDLELPSWLTGNYRIESVLSSSG--KRLGCI 184
Db 228 CSPKR---PMLQRPYFMSFHVGVSVSSSGINKQTGAV 264

RESULT 6

US-10-369-493-20746
; Sequence 20746, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20746
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-20746

Query Match 9.5%; Score 96.5; DB 14; Length 641;
Best Local Similarity 25.0%; Pred. No. 0.64;
Matches 50; Conservative 32; Mismatches 67; Indels 51; Gaps 12;

QY 36 WDCDEGK-----DPAVIRSLTEPDPPIVPGNVTLVSVVSTVPLS----- 77
Db 336 WEVIDKHKNIFYTAPTATRAIMQAGDEPFVKTSKLSURLLGSGVEPINPEAWYHVRV 395

QY 78 ----SPLKVDLVLEKEVAGLWIKIPCTDYI--GSCTFEHFCVDLMDLIP-----T 121
Db 396 GEDRCPI-VDTWQWCTGTGIIITLPGATKLKPSATRPFGVPEIILDPGCVLEGCT 454

QY 122 GEPCC-----PEPLRT-YGLPCHCPKECTYSLPKSEFVV-----PDLEPLGWLATTGNYRI 170
Db 455 GNLCLARSWPQMRVTYG--DHARFEQYFSAIKGYFTGDCRRDITDGFYWITG---RV 509

QY 171 ESVLSSSGKRLGCIKIAASL 190

Db 510 DDVINVGHRMGTAEVESL 529

RESULT 7

US-10-425-115-251307
; Sequence 251307, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants
 ; FILE REFERENCE: 38-21(53222)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369326
 ; SEQ ID NO 251307
 ; LENGTH: 196
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: MRT4577_160777C.1.pep
 US-10-425-115-251307

Query Match 8.7%; Score 88.5; DB 17; Length 196;
 Best Local Similarity 19.1%; Pred. No. 0.92;
 Matches 41; Conservative 37; Mismatches 74; Indels 63; Gaps 8;
 QY 4 LMOAPLIAL--GILLATPA-----CAHLK-----KPSQLS 32
 DB 1 MLRSTLLALTSTLALASVLEPRGLQANAEQVVLGVWGANKLSHVGTDDAGQVGTLT 60
 QY 33 SFWDNDCDEKDPVIRSLTEPPIVVPNGVTLVSVSTSVPLSSPLKVDLVLEKSVAG 92
 DB 61 KWDWDCGSSDALQIDSIKISPPKPGQDLTIVASGRAGSKIDFGTYADYVTKGLIK 120
 QY 93 LWIKIPCTYIGSTCFHFCDVLDMLIPTGEPCEPLRTYGLPCHCPKFGTGYSLPKSEF 152
 DB 121 LLTK-----TFD-VCDELNANAT-----LRCPAPGTHSITQT-- 153
 QY 153 VVPLELPSWLTGNYRIES-VLSSSGKRLGCIKI 186
 DB 154 ----VALPREIPRAKFQVDALVYTQDEEPAACINL 184

RESULT 8
 US-10-437-963-149516
 ; Sequence 149516, Application US/10437963
 ; Publication No. US2004012343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 149516
 ; LENGTH: 152
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_49841C.1.pep
 US-10-437-963-149516

Query Match 8.6%; Score 87.5; DB 16; Length 152;
 Best Local Similarity 23.8%; Pred. No. 0.83;
 Matches 44; Conservative 25; Mismatches 65; Indels 51; Gaps 10;
 QY 9 LLIALGILLATPAQAHLKPKSLSFSDNDCDEKDPVIRSLTEPPIVVPNGVTLV 67
 DB 7 LLLALLAANAANA-----SAVTDEY--CNKGGKYPVKVGVEIYVDPVARGEPATFK 59
 QY 68 VVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTYIGSTCFHFCDVLDMLIPTGEPCE 127
 DB 60 ISASTDKTIG--RGKLVIDVKYFFP-----YVHSETRE-LCDVTS----- 96

QY 128 PLRTYGLPCHCPKFGTGYSLPKSEFVVPDLE-LPSWLTGNYRIE-SVLSSSGKRLGCIK 185
 DB 97 -----CP-----ASGDFLVAHQOITLPSYTPPGSYTITMKMLGDNDELSGIS 138
 QY 186 IAASL 190
 DB 139 FGFSI 143

RESULT 9
 US-09-860-793-3
 ; Sequence 3, Application US/09860793
 ; Patent No. US20020136734A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pruett, John H
 ; APPLICANT: Temeyer, Kevin B
 ; APPLICANT: Kunz, Sidney E
 ; APPLICANT: Fisher, William F
 ; TITLE OF INVENTION: Vaccines for the Protection of Cattle from Psoroptic
 ; TITLE OF INVENTION: Scabies
 ; FILE REFERENCE: Docket 0047.96 - John H. Pruett et al.
 ; CURRENT APPLICATION NUMBER: US/09/860,793
 ; CURRENT FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: 09/366,603
 ; PRIOR FILING DATE: 1999-08-03
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 143
 ; TYPE: PRT
 ; ORGANISM: Psoroptes ovis
 US-09-860-793-3

Query Match 8.4%; Score 86; DB 9; Length 143;
 Best Local Similarity 22.4%; Pred. No. 1.1;
 Matches 44; Conservative 30; Mismatches 50; Indels 72; Gaps 11;
 QY 10 LIALGILLATPAQAHLKPKSLSFSDNDCDEKDPVIRSLTE---PDIYVPGNVT 66
 DB 5 LVWLAITLAVVSAGKVK-----FQDCGKG-----VESLEVEGSGDYCVIHKGKL 51
 QY 67 SV-VGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTYIGSTCFHFCDVLDMLIPTGPC 125
 DB 52 DLASVTSNQDSANLKLIV--ADINGQVLEVPVGDHGD----- 88
 QY 126 PEPLRTYGLPCH---CPFKEG-----TSLPKSEFVVPDLELPSWLTGNYRIESVLS 175
 DB 89 -----CHYVKCPKIKKGQHFVDVKTYSIPA---ILP-----TTKAKIIAKIIG 127
 QY 176 SSGKRLGCIKIAASLK 191
 DB 128 DKGLG-GCIVINGEIQ 142

RESULT 10
 US-10-369-493-20109
 ; Sequence 20109, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374


```

; SEQ ID NO 20109
; LENGTH: 651
; TYPE: PRT
; ORGANISM: No. US2003033675A1toc punctiforme
US-10-369-493-20109

Query Match
Best Local Similarity 8.4%; Score 86; DB 14; Length 651;
Matches 40; Conservative 22; Mismatches 51; Indels 38; Gaps 9;

QY 68 VVGSTSVLSPKVDLVLEKEVAGLMI-----KIPCTDYIGSCTFEH---FCDVLDL-- 117
Db 404 VIGDRCP-----VDTWQTEGGMITPLPCAIPKTP--GSATLPPPGIADVVLDLG 456
QY 118 -LIPTGE-----PCPEPLT-YGLP-----CHCPFKGYSLPKSEFVVPDL 159
Db 457 NTVPNNEGGLAVRHPFPGMMETVYGDPERFRRTYWEHIPPQDGKTYTPAGDGAQD 516
QY 160 PSLWTTGNVRIEVSLSGKRLGCIKIAASL 190
Db 517 YFWMG---RVDDVLNVSGHRLGTMEVESAL 544

RESULT 11
US-10-725-013-2
; Sequence 2, Application US/10725013
; Publication No. US20040198683A1
; GENERAL INFORMATION:
; APPLICANT: Sehgal, Lakshman R.
; TITLE OF INVENTION: Ex vivo and in vivo expression of the thrombomodulin gene
; FILE REFERENCE: 3840-005-27
; CURRENT APPLICATION NUMBER: US/10/725,013
; CURRENT FILING DATE: 2003-12-02
; PRIOR APPLICATION NUMBER: US 60/430,099
; PRIOR FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-725-013-2

Query Match
Best Local Similarity 8.4%; Score 85.5; DB 17; Length 574;
Matches 51; Conservative 21; Mismatches 71; Indels 69; Gaps 13;

QY 8 PLLIALGLLLAT-----PAAHLKKPSQLSFSWNCDEGKDPVIRSLTLEPDP 59
Db 135 PLCVAVSAAEATVPSEPIWEQCEVKADGFLCEPHF-----PATCRPLAVEPG 186
QY 60 VPGNVTLVS-----VGSTSVPLSS-----PLKVDLVL-----EKEVAGL 95
Db 187 AAVSITYGTTPFAARGAQFALPVGSSA--VAPLGQLMCTAPPGAVOGHWAREAPG 243
QY 96 KIPCTDYTGCTFEHFDVLDMLIPTGBCPP-----LRTYGLPCHCPFKGYSLPK-- 149
Db 244 --DCSVENGCC--EACNA-----IFGARPCQCPAGALQADGRSC---TASTQSC 291
QY 150 SEFVVPDLPLPSWLTGNYRIEVSLSGKRL 181
Db 292 EHFVFNPDQP-----GSY---SCMCETGYRL 315

RESULT 12
US-10-094-886-196
; Sequence 196, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Tchernev, Vellizar T.

```

```

; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Vernet, Corine A.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Millier, Charles
; APPLICANT: Casman, Stacie
; APPLICANT: Pena, Carol
; APPLICANT: Gangolli, Esha
; APPLICANT: Gusev, Vladimir
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Gerlach, Valerie
; APPLICANT: Pochart, Pascal
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard
; APPLICANT: Rastelli, Luca
; APPLICANT: Spaderina, Steven
; APPLICANT: LaRochele, William
; APPLICANT: Zhong, Mei
; FILE REFERENCE: 21402-290 B
; CURRENT APPLICATION NUMBER: US/10/094,886
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 196
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-196

Query Match
Best Local Similarity 8.3%; Score 85; DB 15; Length 575;
Matches 49; Conservative 23; Mismatches 76; Indels 60; Gaps 12;

QY 8 PLLIALGLLLAT-----PAAHLKKPSQLSFSWNCDEGKDPVIRSLTLEPDP 59
Db 135 PLCVAVSAAEATVPSEPIWEQCEVKADGFLCEPHF-----PATCRPLAVEPG 186
QY 60 VPGNVTLVS-----VGSTSVPLSS-----PLKVDLVL-----EKEVAGL 97
Db 187 AAVSITYGTTPFAARGAQFALPVGSSA--VAPLGQLMCTAPPGAVOGHWAREAPG 243

```

RESULT 14
 US-10-425-115-279555
 ; Sequence 279555, Application US/10425115
 ; Publication No. US20040214272A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants
 ; FILE REFERENCE: 38-21(53222)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369126
 ; SEQ ID NO 279555
 ; LENGTH: 247
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:

Query Match	8.3%	Score 84.5	DB 17	Length 342
Best Local Similarity	36.5%	Pred. No. 5		
Matches 23	Conservative 6	Mismatches 23	Indels 11	Gaps 3
Qy	124	PCPEPLRTYGLPCPCPFKEGTYSLPKSFVVPDLELP---	SWLTGNTYRIESVLSSSGKR	180
Db	199	PCPDENATGLPFHCDRNLTLPLPS---	MVPGLEAVYRGDWI----	RVEEVPGAFVYN 250
Qy	181	LGC 183		
Db	251	FGC 253		

Search completed: November 17, 2004, 07:01:26
Job time : 27.0941 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 05:43:48 ; Search time 5.41009 Seconds
(without alignments)
3432.447 Million cell updates/sec

Title: US-10-030-937-8

Perfect score: 1018

Sequence: 1 MQSLMQAPLLTALGILLATP.....LSSSGKRLGCIKIASLXGI 193

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 segs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1018	100.0	193	2	154178
2	1009	99.1	200	2	S22411
3	857	84.2	162	2	S13195
4	747.5	73.4	193	2	S35613
5	122	12.0	152	2	T49126
6	90	8.8	456	2	C71402
7	89	8.7	294	2	T34048
8	86	8.4	20	2	S56005
9	85	8.3	171	2	G72563
10	84.5	8.3	1060	2	A10201
11	83	8.2	819	2	B87386
12	81	8.0	353	2	T49440
13	81	8.0	575	1	THHUB
14	80	7.9	551	2	AC2311
15	78.5	7.7	388	2	F71029
16	78	7.7	467	2	S19317
17	78	7.7	715	2	D96582
18	77.5	7.6	249	2	A99222
19	77.5	7.6	693	2	T25878
20	77	7.6	432	2	A25483
21	77	7.6	463	2	T21042
22	77	7.6	1642	2	T19130
23	76.5	7.5	504	2	AD3629
24	76.5	7.5	583	2	F69153
25	76.5	7.5	2482	2	I48922
26	76.5	7.5	2483	1	A49617
27	76	7.5	785	2	A29953
28	75.5	7.4	149	2	I69229
29	75.5	7.4	271	1	PKCT9

ALIGNMENTS

RESULT 1

I54178

ganglioside M2 activator protein precursor - human

C:Species: Homo sapiens (man)

C:Date: 24-May-1996 #sequence revision 24-May-1996 #text_change 09-Jul-2004

C:Accession: I54178; JQ1037; S05036; S22410; S17107

R:Xi'e, B.; Kennedy, J.L.; McInnes, B.; Auger, D.; Mahuran, D.

Genomics 14, 796-798, 1992

A:Title: Identification of a processed pseudogene related to the functional gene encoding man chromosome 5.

A:Reference number: I54178; MUID:93052421; PMID:1427911

A:Accession: I54178

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-193 <RES>

A:Cross-references: UNIPROT:P17900; GB:L01439; NID:gl83358; PIDN:AAA52767.1; PID:gl83359

R:Xi'e, B.; McInnes, B.; Neote, K.; Lamhonwah, A.M.; Mahuran, D.

Biochem. Biophys. Res. Commun. 177, 1217-1223, 1991

A:Title: Isolation and expression of a full-length cDNA encoding the human GM2 activator

A:Reference number: JQ1037; MUID:91282768; PMID:2059210

A:Accession: JQ1037

A:Molecule type: mRNA

A:Residues: 1-18, 'A', '20-193 <XIE>

A:Cross-references: GB:M76477; NID:gl83356; PIDN:AAA35907.1; PID:gl83357

A:Experimental source: HeLa cell

A:Note: 19-Thr and 69-Met were also found

R:Schroeder, M.; Klima, H.; Nakano, T.; Kwon, H.; Quintern, L.E.; Gaertner, S.; Suzuki, F.

FEBS Lett. 251, 197-200, 1989

A:Title: Isolation of a cDNA encoding the human G(M2) activator protein.

A:Reference number: S05036; MUID:89325664; PMID:2753159

A:Accession: S05036

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 15-193 <SCH>

A:Cross-references: GB:X16087; NID:gl81852; PIDN:CAA34215.1; PID:gl81853

R:Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.

Biochem. J. 282, 807-813, 1992

A:Title: Evidence for two cDNA clones encoding human GM2-activator protein.

A:Reference number: S22410; MUID:92207171; PMID:1554364

A:Accession: S22410

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 43-142, 'VST', <NAG>

A:Cross-references: EMBL:X61094; NID:gl81854; PIDN:CAA43407.1; PID:gl81855

C:Comment: This protein transports GM2 ganglioside from the lysosomal membrane to hexosan

C:Genetics:

A:Gene: GDB:GM2A

A:Cross-references: GDB:120000; OMIM:272750

A:Map position: 5q32-5q33

F:1-23/Domain: signal sequence #status predicted <SIG>

F:32-193/Product: GM2 ganglioside activator protein #status predicted <GM2>

F:32-193/Product: ganglioside M2 activator #status predicted <MAT>

angiotensin precu
hypothetical prote
protein T22A3.8 [i
hypothetical prote
hypothetical prote
laminin alpha chai
alpha-1 proteinase
APH protein - Salm
dimethyladenosine
probable decarboxy
probable glutamate
acetyl coenzyme A
hypothetical prote
genome polyprotein
protein UNC-89 - C
alpha-2 macroglobu
hypothetical prote

Query Match 100.0%; Score 1018; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 3.4e-86;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIALGLLLATPAQAHLKPSQLSFSWNCDEGKDPVIRSLTLEPPIV 60
DB 1 MQSLMQAPLLIALGLLLATPAQAHLKPSQLSFSWNCDEGKDPVIRSLTLEPPIV 60

QY 61 PGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
DB 61 PGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120

QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180
DB 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180

QY 181 LGCIIKIAASLKI 193
DB 181 LGCIIKIAASLKI 193

RESULT 2
S22411
ganglioside M2 activator protein (clone pGAP2) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
C;Accession: S22411; S17108
R;Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A;Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A;Reference number: S22410; MUID:92207171; PMID:1554364
A;Accession: S22411
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-200 <NAG>
A;Cross-references: UNIPROT:PI7900; EMBL:X61095; NID:g31856; PIDN:CAAM3408.1; PID:g31857

Query Match 99.1%; Score 1009; DB 2; Length 200;
Best Local Similarity 98.4%; Pred. No. 2.4e-85;
Matches 190; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIALGLLLATPAQAHLKPSQLSFSWNCDEGKDPVIRSLTLEPPIV 60
DB 8 MQSLMQAPLLIALGLLLATPAQAHLKPSQLSFSWNCDEGKDPVIRSLTLEPPIV 67

QY 61 PGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 120
DB 68 PGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 127

QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 180
DB 128 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 187

QY 181 LGCIIKIAASLKI 193
DB 188 LGCIIKIAASLKI 200

RESULT 3
S13195
ganglioside M2 activator protein - human
C;Species: Homo sapiens (man)
C;Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S13195
R;Fuerst, W.; Schubert, J.; Machleidt, W.; Meyer, H.E.; Sandhoff, K.
Eur. J. Biochem. 192, 709-714, 1990
A;Title: The complete amino-acid sequences of human ganglioside GM2 activator protein and
A;Reference number: S13195; MUID:91006165; PMID:2209618
A;Accession: S13195
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-162 <FUE>

A;Cross-references: UNIPROT:PI7900

Query Match 84.2%; Score 857; DB 2; Length 162;
Best Local Similarity 99.4%; Pred. No. 1.7e-71;
Matches 159; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 32 SSFSDNCDEGKDPVIRSLTLEPPIVPGNVTLISVVGSTSVPLSSPLKVDLVLEKEVA 91
DB 1 SSFSDNCDEGKDPVIRSLTLEPPIVPGNVTLISVVGSTSVPLSSPLKVDLVLEKEVA 60

QY 92 GLWIKIPCTDYIGSCTFEHFCVDLMDLIPGTGPPCEPLRTYGLPCHCPKFGTYSLPKSE 151
DB 61 GLWIKIPCTDYIGSCTFEHFCVDLMDLIPGTGPPCEPLRTYGLPCHCPKFGTYSLPKSE 120

QY 152 FVVPDLELPSWLTGNYRIESVLSGSKLGCIIKIAASLK 191
DB 121 FVVPDLELPSWLTGNYRIESVLSGSKLGCIIKIAASLK 160

RESULT 4
S35613
ganglioside M2 activator protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 07-May-1999
C;Accession: S35613
R;Bellachio, G.; Stirling, J.L.; Orlacchio, A.; Beccari, T.
Biochem. J. 294, 227-230, 1993
A;Title: Cloning and sequence analysis of a cDNA clone coding for the mouse GM2
A;Reference number: S35613; MUID:93371367; PMID:7689829
A;Accession: S35613
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-193 <BEL>

Query Match 73.4%; Score 747.5; DB 2; Length 193;
Best Local Similarity 72.3%; Pred. No. 2.4e-61;
Matches 136; Conservative 23; Mismatches 26; Indels 3; Gaps 1;

QY 8 PLLIALGLLLA---TQAQHLKPSQLSFSWNCDEGKDPVIRSLTLEPPIVPGNV 64
DB 5 PLLIALGLLLAGSAPARLVKPSQLSGLGFSWNCDEGKDPVIRSLTLEPPIVPGNV 64

QY 65 TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP 124
DB 65 VLSLEKGTSVPLTAPQKVELTVEKAVAGFWIKIPCVQLGSCVENICDLIDEYIPGES 124

QY 125 CPPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKRLGCI 184
DB 125 CPPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKRLGCI 184

QY 185 KIAASLKG 192
DB 185 KIAASLKG 192

RESULT 5
T49126
hypothetical protein F26G5.50 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C;Accession: T49126
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.;
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25017
A;Accession: T49126
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-152 <DAN>
A;Cross-references: UNIPROT:Q9LXQ2; EMBL:AL353814; GSPDB:GN00061; ATSP:F26G5.50
C;Genetics:
A;Gene: ATSP:F26G5.50
A;Map position: 3

T34048
hypothetical protein W08P4.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*

C;Accession: G72563

C;Accession: G72563

C;Access

R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
 A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
 A;Reference number: A72450; MUID:99310339; PMID:10382966
 A;Accession: G72563
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-171 <RAW>
 A;Cross-references: UNIPROT:Q9YB02; DBJ:AF000062; NID:95105244; PIDN:BAA80796.1; PID:dl
 A;Experimental source: strain K1
 C;Genetics:
 A;Gene: APE1793

Query Match 8.3%; Score 85; DB 2; Length 171;
 Best Local Similarity 25.0%; Pred. No. 1.5;
 Matches 31; Conservative 17; Mismatches 44; Indels 32; Gaps 4;
 QY 10 LIALGILLATPAQAHLKPKSLSFS-----WDNCDEKDPVIRSLT 52
 Db 3 MCLSLILETSLTPTSTPISLSFTRAGRFLAPLILASVTWGNCSFGKPLMVLRLIS 62
 QY 53 L-----BPDPIVPGNVT--LSVVGSTSVPLSKVDLVLEKEVAGLWIKIPCTDYI 103
 Db 63 TSSLILRSSPAPPTTHTTLGLSLGKAPSLSSILNLSLIDAN-----RSPISAAT 116
 QY 104 GSCT 107
 Db 117 GSCT 120

RESULT 10
 AI0201
 beta-galactosidase (EC 3.2.1.23) [imported] - Yersinia pestis (strain CO92)
 C;Species: Yersinia pestis
 C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
 C;Accession: AI0201
 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001
 A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A;Reference number: AB0001; MUID:21470413; PMID:11586360
 A;Accession: AI0201
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1060 <KUR>
 A;Cross-references: UNIPROT:Q8ZFP0; GB:AL590842; PIDN:CAC90476.1; PID:g15979691; GSPDB:G
 A;Gene: lacZ
 C;Superfamily: beta-galactosidase
 C;Keywords: glycosidase; hydrolase

Query Match 8.3%; Score 84.5; DB 2; Length 1060;
 Best Local Similarity 24.2%; Pred. No. 15;
 Matches 43; Conservative 18; Mismatches 54; Indels 63; Gaps 10;
 QY 20 PQAHLKPKSLSFSWNCDEKDPVIRSLTLEPDPVPGNVTLSVVGSTSVPLSSP 79
 Db 8 PLQVQLSLPQLSLRRDWN-----PQITQYHRLAHP--PFH-SWRDVESAQRDRSP 57
 QY 80 LKVDLVLEKEVAGLWIKIPCTDYIGSCTFEH-----CDVLDMLIPTGCPCE 127
 Db 58 -----QQQTNLGLW-----SPSYTQPAVEHVRCDLAE-----AKPLPV 94
 QY 128 P-----LRTYGLCHCFKKGTYSLPKSEFVVPDL-----LPSWLTATGNRYI 170
 Db 95 PANWQLHGVDABIYTNIQ---YPIPNPRVFDLNPFGCVSRDFTLEPSKSLASGKTRI 149

RESULT 11
 B87386
 hypothetical protein CC1102 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus
 C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
 C;Accession: B87386
 R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J. B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A;Title: Complete Genome Sequence of Caulobacter crescentus.
 A;Reference number: A87249; MUID:21173698; PMID:11259647
 A;Accession: B87386
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-819 <STO>
 A;Cross-references: UNIPROT:Q9A992; GB:AE005673; NID:gl3422408; PIDN:AAK23086.1; GSPDB: B87386
 C;Genetics:
 A;Gene: CC1102

Query Match 8.2%; Score 83; DB 2; Length 819;
 Best Local Similarity 23.7%; Pred. No. 16;
 Matches 40; Conservative 21; Mismatches 46; Indels 62; Gaps 10;
 QY 36 WDN-----CDEGKDPVIRSLTLEP-----DPIVPGNVT-LSVVGSTSVPLS 77
 Db 652 WDRSAYIRQDRDPTLVSMASRPGEVILWAGDVEPWLVRASWISKVQSGVFS 711
 QY 78 SPUKV---DLVLEKEVAGL---WIKIPCTDYIGSCTFEHFCVDLMLIPTGCPCEPLR 130
 Db 712 RPLAMALHDRVARVEAAGLAGPDWIR-PL-----LMDALKPP-PPKPEKVR 755
 QY 131 TYGLPCHCFKKGTYSLPKSEFVVPDLPLSPWLTGNYRIESVLSGSK 179
 Db 756 AF-----CAASDA-----PAMIVSPLWDGEGVLDAGLK 783

RESULT 12
 T49440
 alcohol dehydrogenase (EC 1.1.1.1) I adh1 B17C10.210 [similarity] - Neurospora crassa
 N;Alternate names: protein B17C10.210
 C;Species: Neurospora crassa
 C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
 C;Accession: T49440
 R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, May 2000
 A;Reference number: Z25022
 A;Accession: T49440
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-353 <SCH>
 A;Cross-references: UNIPROT:Q9P6C8; EMBL:AL355926; GSPDB:GN00116; NCSP:B17C10.210
 A;Experimental source: BAC clone B17C10; strain OK74A
 C;Genetics:
 A;Gene: NCSP:B17C10.210

A;Map position: 6
 A;Introns: 21/1
 C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C;Keywords: NAD; oxidoreductase; zinc
 F;47,70,157/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
 Query Match 8.0%; Score 81; DB 2; Length 353;
 Best Local Similarity 27.3%; Pred. No. 8.6;
 Matches 53; Conservative 26; Mismatches 61; Indels 54; Gaps 16;
 QY 36 WNCDEGK-DPAVIRSLTLE---PDPVPGNVTLSVVGSTSV-----PLSS--PL-- 80
 Db 11 MAQVVEKGGPVVFQIPVQKPGPDEVLI--NVKSGVCHTDLHAMKGDWPLATKMLPLVG 68
 QY 81 -----KVDLVLEKEV---AGL-WIKIPCTDYIGSCTFEHFCVDLMLIPTGEP-C 125
 Db 69 GHEGAGVVVAKGELVTEVEVGHGKMLN-----GSLCACSFQMQAD-----EPLC 115

Query Match 8.0%; Score 81; DB 2; Length 353;
 Best Local Similarity 27.3%; Pred. No. 8.6;
 Matches 53; Conservative 26; Mismatches 61; Indels 54; Gaps 16;
 QY 36 WNCDEGK-DPAVIRSLTLE---PDPVPGNVTLSVVGSTSV-----PLSS--PL-- 80
 Db 11 MAQVVEKGGPVVFQIPVQKPGPDEVLI--NVKSGVCHTDLHAMKGDWPLATKMLPLVG 68
 QY 81 -----KVDLVLEKEV---AGL-WIKIPCTDYIGSCTFEHFCVDLMLIPTGEP-C 125
 Db 69 GHEGAGVVVAKGELVTEVEVGHGKMLN-----GSLCACSFQMQAD-----EPLC 115
 QY 126 PEPLRTYGLPCHCFKKGTYSLPKSEFV--VF-----DLELPSWLTGNYRIESVLSGSKR 180
 Db 116 PHALLS-GYTVDSGSPQ--YATAKAHVAKIPKGCDELTTAPVLCAGITVYKLGESGVR 172

A:Reference number: A52804; PDB:1HLT

C;Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synthetic protein AII4042 (imported)

C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Accession: AC2311
#sequence_revision 14 DEC-2001 #text_change 09-001-2004

R; Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, H.

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena cylindrica

A;Accession: AC2311
A-Status: preliminary

A:Molecule type: DNA
A:Residues: 1-551 <K1R>

A;Experimental source: strain PCC 7120

A;Gene: all4042

Query Match 7.9%; Score 80; DB 2; Length 551;

Matches 51; Conservative 36; Mismatches 79; Indels 60; Gaps 12;

[illegible]

(CZ) TREATMENT IN THE COUNTRIES OF ORIGIN AND DESTINATION

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Variable	Mean	Standard Deviation	Minimum	Maximum
Age	35.2	12.5	22	65
Gender	0.48	0.50	0	1
Marital Status	0.65	0.48	0	1
Education	12.8	2.1	9	16
Income	45000	15000	20000	80000
Health	0.72	0.45	0	1
Employment	0.85	0.35	0	1
Home Ownership	0.55	0.50	0	1
Vehicle Ownership	0.68	0.47	0	1
Life Satisfaction	4.2	1.8	1	7
Life Satisfaction (Control)	4.1	1.7	1	7

ZZ ZYXWVUTS--DITKAT--VFADVNQNZGI---NLGQDIFSMVQGEIAY 364

QY 155 PDL-----ELPSWLTIGNY-----RIESVLSGKRLGCIKI 186

Db 365 ALIPRTGQSI PDWV FVTEKSEN VPEAIGRLDAIASSQGLSTNTIKL 410

RECORD IS
F71029

C:Species: *Pyrococcus horikoshii*
hypothetical protein PH1528 - *Pyrococcus horikoshii*

C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004

R; Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin

III, CHIRAKA, I.; ICHIMATSU, I.; ICHIKAWA, I.; NUGOLI, J.; KUSHIDA, N.; OGUCHI,
 DNA Res. 5, 55-76, 1998

A;11c1c: Complete sequence and gene organization of the genome of a hyper-thermophilic A

A;Accession: F71029

A;Molecule type: DNA

A;Cross-references: UNIPROT:O59197; GB:AP000006; NID:q3236133; PIDN:BAA30638.1; PID:q3236133

A/Experimental source: strain 013

A/Note: this accession replaces an interim accession for a sequence replaced by GenBank

C;Genetics:
A.Gene: PH1538

Best Local Similarity 26.0%; Pred. No. 16;

matches 30; conservative 23; mismatches 34; indels 31; gaps 8;

QY 13 LGLLATPA---QAHLKKPSQLSSFSWDNCDEG--KDPAVIRS-----LTLEPDPVW 60

db 211 LDRVIITPPVHDAHI---DELDSFSFQEEVENVEPPLFVKAMKGLVLTEPPRLVV 267

QY 61 -----PCNVTLSVVG-STSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFC 112

Db 268 MLLKLRGEVKVSLKGRQIRIPLNFTLVVDTKPERYSGL--KFPPIRINLPFFDDETF 325

113 DVI DML TPTCED-----CDEPIBTV 133

[illegible]

100

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 05:42:53 ; Search time 29.6589 Seconds
(without alignments)
3744.152 Million cell updates/sec

Title: US-10-030-937-8

Perfect score: 1018

Sequence: 1 MQSLMQAPLLIALGLLALTP.....LSSSGKRLGCIKIAASLKI 193

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1018	100.0	193	1	SAP3 HUMAN
2	1000	98.2	189	2	Q8LBI5
3	1000	98.2	189	2	Caa43994
4	920.5	90.4	190	2	Q8HXX6
5	748.5	73.5	199	2	Q6IN37
6	748.5	73.5	199	2	Q8CJH4
7	746.5	73.3	193	1	SAP3 MOUSE
8	602	59.1	151	2	Q6Q7X5
9	602	59.1	151	2	AAS64350
10	544	53.4	103	2	Q14427
11	519	51.0	146	2	Q6Q7X4
12	519	51.0	146	2	AAS64351
13	512.5	50.3	197	2	Q6GLN6
14	262	25.7	217	2	Q75R48
15	262	25.7	217	2	BAD16601
16	231	22.7	47	2	Q8I028
17	138	13.6	40	2	Q29283
18	123	12.1	242	2	Q6J1S7
19	123	12.1	242	2	AAT35232
20	122	12.0	152	2	Q8LXQ2
21	111	10.9	26	2	Q8I029
22	108.5	10.7	155	2	Q8I6V6
23	107.5	10.6	243	2	Q7S5Y5
24	107.5	10.6	331	2	Q6NE93
25	107.5	10.6	331	2	CAF06269
26	101	9.9	480	2	Q8VQNI
27	98	9.6	480	2	Q700B8
28	98	9.6	480	2	Q9WXD3
29	98	9.6	480	2	CAF74787
30	97	9.5	556	2	Q87548
31	96.5	9.5	650	2	Q6ND98

32	96.5	9.5	650	2	CAE25655
33	93.5	9.2	323	2	Q8EL09
34	92	9.0	373	2	Q8JVA9
35	91	8.9	160	2	Q8L8M3
36	91	8.9	160	2	Q9AST8
37	90	8.8	456	2	Q23270
38	90	8.8	466	2	Q8RY86
39	89	8.7	294	2	Q02106
40	88	8.6	339	2	Q8QQX8
41	87.5	8.6	152	2	Q8H559
42	86.5	8.5	642	2	Q8GMD9
43	86	8.4	20	2	Q9QUW2
44	86	8.4	143	1	ALL2_PSOOV
45	85	8.3	171	2	Q9YB02

ALIGNMENTS

RESULT 1

ID	SAP3 HUMAN	STANDARD;	PRT;	193 AA.
AC	P17900; Q14426; Q14428;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	01-OCT-2004 (Rel. 45, Last annotation update)			
DE	Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate activator protein) (Shingolipid activator protein 3) (SAP-3).			
GN	Name=GM2A;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A., AND VARIANT ALA-19.			
RX	MEDLINE=91282768; PubMed=2059210;			
RA	Xie B., McInnes B., Neote K., Lamhonwah A.-M., Mahuran D.;			
RT	"Isolation and expression of a full-length cDNA encoding the human G-2 activator protein."			
RL	Biochem. Biophys. Res. Commun. 177:1217-1223(1991).			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92008637; PubMed=1915857;			
RA	Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K., Sandhoff K.;			
RT	"Characterization of full-length cDNAs and the gene coding for the human GM2 activator protein."			
RL	FEBS Lett. 289:260-264(1991).			
RP	SEQUENCE FROM N.A., AND VARIANT ALA-19.			
RC	TISSUE=Placenta;			
RX	MEDLINE=92207171; PubMed=1554364;			
RA	Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;			
RT	"Evidence for two cDNAs encoding human GM2-activator protein.";			
RL	Biochem. J. 282:807-813(1992).			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93052421; PubMed=1427911;			
RA	Xie B., Kennedy J.L., McInnes B., Auger D., Mahuran D.J.;			
RT	"Identification of a processed pseudogene related to the functional gene encoding the GM2 activator protein: localization of the pseudogene to human chromosome 3 and the functional gene to human chromosome 5.";			
RL	Genomics 14:796-798(1992).			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=9294584; PubMed=10364519;			
RA	Chen B., Rigat B., Curry C., Mahuran D.J.;			
RT	"Structure of the GM2A gene: identification of an exon 2 nonsense mutation and a naturally occurring transcript with an in-frame deletion of exon 2.";			
RL	Am. J. Hum. Genet. 65:77-87(1999).			

Cae25655	rhodopsin
Q8el09	oceanobacil
Q8jva9	tioman viru
Q8l8m3	arabidopsis
Q9ast8	arabidopsis
Q23270	arabidopsis
Q8ry86	arabidopsis
Q02106	caenorhabd
Q8qgx8	meleagrid h
Q8h559	oryza sativ
Q8gmd9	streptomyce
Q9quw2	rattus sp.
Q9fse2	psoroptes o
Q9ybo2	aeropyrum p

RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Gaertner S., Suzuki K., Sandhoff K.;
RA "Isolation of a cDNA encoding the human GM2 activator protein.";
RL FEBS Lett. 251:197-200(1989).
RN [8]
RP SEQUENCE OF 32-193.
RC TISSUE=Kidney;
RX MEDLINE=9325664; PubMed=2753159;
RA Schroeder M., Klima H., Nakano T., Kwon H., Quintern L.B.,
RA Gaertner S., Suzuki K., Sandhoff K.;
RT "Isolation of a cDNA encoding the human GM2 activator protein.";
RL FEBS Lett. 251:197-200(1989).
RN [8]
RP SEQUENCE OF 32-193.
RC TISSUE=Kidney;
RX MEDLINE=9325664; PubMed=2753159;
RA Schroeder M., Klima H., Nakano T., Kwon H., Quintern L.B.,
RA Gaertner S., Suzuki K., Sandhoff K.;
RT "Isolation of a cDNA encoding the human GM2 activator protein.";
RL FEBS Lett. 251:197-200(1989).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.0 ÅNGSTROMS).
RX MEDLINE=20545600; PubMed=11090283; DOI=10.1006/jmbi.2000.4225;
RA Wright C.S., Li S.-C., Rastinejad F.;
RT "Crystal structure of human GM2-activator protein with a novel beta-
cup topology.";
RL J. Mol. Biol. 304:411-422(2000).
RN [10]
RP VARIANT TSD-AB ARG-138.
RX MEDLINE=92008638; PubMed=1915858;
RA Schroeder M., Schnabel D., Hurwitz R., Young E., Suzuki K.,
RT "A mutation in the gene of a glycolipid-binding protein (GM2
activator) that causes GM2-gangliosidosis variant AB.";
RL FEBS Lett. 290:1-3(1991).
RN [11]
RP VARIANT TSD-AB PRO-169.
RX MEDLINE=94063850; PubMed=8244332;
RA Schroeder M., Schnabel D., Hurwitz R., Young E., Suzuki K.,
RT "Molecular genetics of GM2-gangliosidosis AB variant: a novel mutation
and expression in BHK cells.";
RL Hum. Genet. 92:437-440(1993).
RN [12]
RP VARIANT TSD-AB LYS-88 DEL.
RX MEDLINE=97055887; PubMed=8900233;
RA Schepers U., Glombitza G., Lemm T., Hoffmann A., Chabas A., Ozand P.,
RA Sandhoff K.;
RT "Molecular analysis of a GM2-activator deficiency in two patients with
GM2-gangliosidosis AB variant.";
RL Am. J. Hum. Genet. 59:1048-1056(1996).
CC -|- FUNCTION: Binds gangliosides and stimulates ganglioside GM2
degradation. It stimulates only the breakdown of ganglioside GM2
and glycolipid GA2 by beta-hexosaminidase A. It extracts single
GM2 molecules from membranes and presents them in soluble form to
beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and

conversion to GM3.
-|- SUBCELLULAR LOCATION: Lysosomal.
-|- DISEASE: Defects in GM2A are the cause of Tay-Sachs disease AB
variant (TSD-AB) [MIM:272750]; also known as GM2-gangliosidosis
type AB.
-|- DATABASE: NAME=GM2AB; NOTE=GM2A mutation database;
WWW="http://www.hexdb.mcgill.ca/?Topic=GM2AB&Page=MutationSubmission".

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; M76477; AAA35907.1; --
EMBL; X62078; CAA43993.1; --
EMBL; X61095; CAA43408.1; ALT_INIT.
EMBL; L01433; AAS2767.1; --
EMBL; AF124719; AAD25741.1; --
EMBL; AF124717; AAD25741.1; JOINED.
EMBL; BC009273; AAH09273.1; --
EMBL; X16087; CAA34215.1; --
PIR; I54178; I54178.
PIR; S13195; S13195.
PIR; S22411; S22411.
PDB; 1GI3; X-ray; A/B/C=32-193.
Genew; HGNC:4367; GM2A.
MIM; 272750; --
GO; GO:0005764; C:lysosome; NAS.
GO; GO:0030290; F:sphingolipid activator protein activity; NAS.
GO; GO:0019377; P:glycolipid catabolism; NAS.
GO; GO:0030149; P:sphingolipid catabolism; NAS.
InterPro; IPR003172; E1_DerP2_DerF2.
SMART; SM00737; ML; 1
3D-structure; Direct protein sequencing; Disease mutation; Signal;
Glycoprotein; GM2-gangliosidosis; Lysosome; Polymorphism; Signal;
Sphingolipid metabolism.
FT SIGNAL 1 31
FT CHAIN 32 193 Ganglioside GM2 activator.
FT DISULFID 39 183
FT DISULFID 99 106
FT DISULFID 112 138
FT DISULFID 125 136
FT CARBOHYD 63 63
FT VARIANT 19 19 N-linked (GLCNAC...).
FT T -> A.
FT FTID=VAR_013830.
FT VARIANT 32 33 Missing (in 80% of the protein).
FT FTID=VAR_006946.
FT VARIANT 88 88 Missing (in TSD-AB).
FT C -> R (in TSD-AB).
FT FTID=VAR_011697.
FT R -> P (in TSD-AB).
FT FTID=VAR_006947.
FT FTID=VAR_011698.
FT V -> I (in Ref. 3).
FT V -> M (in Ref. 3).
FT CONFLICT 59 59
FT CONFLICT 69 69
FT STRAND 35 38
FT TURN 41 43
FT STRAND 46 54
FT STRAND 58 59
FT STRAND 63 72
FT STRAND 76 76
FT STRAND 81 90
FT TURN 91 92
FT STRAND 93 96
FT STRAND 100 100
FT TURN 101 102
FT STRAND 103 103
FT STRAND 107 108
FT TURN 109 110

```

FT HELIX      111      118
FT TURN      121      122

Query Match      100.0%; Score 1018; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 2.1e-84;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOSLMQAPLLIALGILLATPAQAHLKPSQLSFSFWDNCEGKDPVIRSLTLEPDPPIV 60
DB 1 MOSLMQAPLLIALGILLATPAQAHLKPSQLSFSFWDNCEGKDPVIRSLTLEPDPPIV 60
QY 61 PGNVTLVVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYGSCTFEHCDFVLDMLIP 120
DB 61 PGNVTLVVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYGSCTFEHCDFVLDMLIP 120
QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGKRLGCI 180
DB 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGKRLGCI 180
QY 181 LGCIIKIAASLKGI 193
DB 181 LGCIIKIAASLKGI 193

RESULT 2
Q6L5L5 PRELIMINARY; PRT; 189 AA.
AC Q6L5L5;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE GM2 activator protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=8503891;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
human GM2 activator protein."
RL FEBS Lett. 289:260-264(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277527; PubMed=8503891;
RA Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA Sandhoff K.;
RT "Over-expression of a functionally active human GM2-activator protein
in escherichia coli."
RL Biochem. J. 292:571-576(1993).
DR EMBL; X62078; CAA43994.1; -.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;
[1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=8503891;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
human GM2 activator protein."
RL FEBS Lett. 289:260-264(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277527; PubMed=8503891;
RA Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA Sandhoff K.;
RT "Over-expression of a functionally active human GM2-activator protein
in escherichia coli."
RL Biochem. J. 292:571-576(1993).
DR EMBL; X62078; CAA43994.1; -.
DR InterPro; IPR003172; E1_DerP2_DerE2.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match      98.2%; Score 1000; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 8.7e-83;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 MQAPLLIALGILLATPAQAHLKPSQLSFSFWDNCEGKDPVIRSLTLEPDPPIVPGNV 64
DB 1 MQAPLLIALGILLATPAQAHLKPSQLSFSFWDNCEGKDPVIRSLTLEPDPPIVPGNV 60
QY 65 TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYGSCTFEHCDFVLDMLIPTGEP 124
DB 61 TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYGSCTFEHCDFVLDMLIPTGEP 120
QY 125 CPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGKRLGCI 184
DB 121 CPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGKRLGCI 180
QY 185 KIAASLKGI 193
DB 181 KIAASLKGI 189

RESULT 4
Q8HXX6 PRELIMINARY; PRT; 190 AA.
AC Q8HXX6;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Ganglioside GM2 activator.
GN Name=gm2a;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]_
RP SEQUENCE FROM N.A.

```

```

QY 185 KIAASLKGI 193
DB 181 KIAASLKGI 189

RESULT 3
CAA43994 PRELIMINARY; PRT; 189 AA.
AC CAA43994;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE GM2 activator protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=8503891;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
human GM2 activator protein."
RL FEBS Lett. 289:260-264(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277527; PubMed=8503891;
RA Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA Sandhoff K.;
RT "Over-expression of a functionally active human GM2-activator protein
in escherichia coli."
RL Biochem. J. 292:571-576(1993).
DR EMBL; X62078; CAA43994.1; -.
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match      98.2%; Score 1000; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 8.7e-83;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 MQAPLLIALGILLATPAQAHLKPSQLSFSFWDNCEGKDPVIRSLTLEPDPPIVPGNV 64
DB 1 MQAPLLIALGILLATPAQAHLKPSQLSFSFWDNCEGKDPVIRSLTLEPDPPIVPGNV 60
QY 65 TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYGSCTFEHCDFVLDMLIPTGEP 124
DB 61 TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYGSCTFEHCDFVLDMLIPTGEP 120
QY 125 CPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGKRLGCI 184
DB 121 CPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGKRLGCI 180
QY 185 KIAASLKGI 193
DB 181 KIAASLKGI 189

RESULT 4
Q8HXX6 PRELIMINARY; PRT; 190 AA.
AC Q8HXX6;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Ganglioside GM2 activator.
GN Name=gm2a;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]_
RP SEQUENCE FROM N.A.

```

```
RC TISSUE=Brain cerebellum cortex;
RA Kusuda J., Oeada N., Hida M., Sugano S., Hashimoto K.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB083313; BAC20592.1; -.
DR HSP; P17900; IG13.
DR InterPro: IPR003172; EI_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 190 AA; 20494 MW; 9F9582BBB75715C3 CRC64;

Query Match 90.4%; Score 920.5; DB 2; Length 190;
Best Local Similarity 90.2%; Pred. No. 1.5e-75;
Matches 174; Conservative 8; Mismatches 8; Indels 3; Gaps 1;

Qy 1 MQSIMGAPVLIAGLLIATPAQAHLKPKSQSLSPSSWNCDEGKDPVIRSLTLEPDPV 60
Db 1 MQSIMGAPVLIAGLLIATPAQAHLKPKSQSLSPSSWNCDEGKDPVIRSLTLEPDP 57
Qy 61 PGNVTVSVGSGTSPVLSPLKVDLVLKEVAGLWIKIPCTDYIGSCTFEHFCVDL 120
Db 58 PGNVTVSVGSGTSPVLSPLKVDLVLKEVAGLWIKIPCTDYIGSCTFEHFCVDL 117
Qy 121 TGBPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNRYIESVL 180
Db 118 TGBPCPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNRYIESVL 177
Qy 181 LGCIKIAASLKG 193
Db 178 LGCIKIAASLKG 190

RESULT 5
Q6IN37 PRELIMINARY; PRT; 199 AA.
AC Q6IN37;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE GM2 ganglioside activator protein.
GN Name=Gm2a;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Stalska U., Smalls D.B., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RP "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC072474; AAH72474.1; -.

DR InterPro: IPR003172; EI_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 199 AA; 21493 MW; C2B5203FDFABF507 CRC64;

Query Match 73.5%; Score 748.5; DB 2; Length 199;
Best Local Similarity 67.7%; Pred. No. 6.6e-60;
Matches 134; Conservative 24; Mismatches 31; Indels 9; Gaps 1;

Qy 4 LMQAPLLIALGLLL-----ATPAQAHLKPKSQSLSPSSWNCDEGKDPVIRSLTLE 54
Db 1 MRRVPLLVLGLLLFVLGLFAGVPVAPSRLLSKRPSQLGGFSDNCDEGKDPVIRSLT 60
Qy 55 PDPVTVGNTVLSVVGSTSPVLSPLKVDLVLKEVAGLWIKIPCTDYIGSCTFEHFCDV 114
Db 61 PDPVTVGNTVLSVVGSTSPVLSPLKVDLVLKEVAGLWIKIPCTDYIGSCTFEHFCDV 120
Qy 115 LDMLIPTGEPCEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNRYIESVL 174
Db 121 IDQYIPGECPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNRYIESVL 180
Qy 175 SSGKRLGCIKIAASLKG 192
Db 181 SSGKRLGCIKIAASLKG 198

RESULT 6
Q8CJH4 PRELIMINARY; PRT; 199 AA.
AC Q8CJH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE GM2 activator protein.
GN Name=rGM2AP;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Miwa N., Okada T., Nakamura S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051391; BAC24018.1; -.
DR HSP; P17900; IG13.
DR InterPro: IPR003172; EI_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 199 AA; D585203FDFABF507 CRC64;

Query Match 73.5%; Score 748.5; DB 2; Length 199;
Best Local Similarity 67.7%; Pred. No. 6.6e-60;
Matches 134; Conservative 24; Mismatches 31; Indels 9; Gaps 1;

Qy 4 LMQAPLLIALGLLL-----ATPAQAHLKPKSQSLSPSSWNCDEGKDPVIRSLTLE 54
Db 1 MRRVPLLVLGLLLFVLGLFAGVPVAPSRLLSKRPSQLGGFSDNCDEGKDPVIRSLT 60
Qy 55 PDPVTVGNTVLSVVGSTSPVLSPLKVDLVLKEVAGLWIKIPCTDYIGSCTFEHFCDV 114
Db 61 PDPVTVGNTVLSVVGSTSPVLSPLKVDLVLKEVAGLWIKIPCTDYIGSCTFEHFCDV 120
Qy 115 LDMLIPTGEPCEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNRYIESVL 174
Db 121 IDQYIPGECPEPLRTYGLPCHCPKFGTYSLPKSEFVVPDLELPSWLTGNRYIESVL 180
Qy 175 SSGKRLGCIKIAASLKG 192
Db 181 SSGKRLGCIKIAASLKG 198

RESULT 7
SAP3 MOUSE
ID SAP3 MOUSE STANDARD; PRT; 193 AA.
AC Q60648; Q61610; Q61819;
```

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate

DE activator protein) (Shingolipid activator protein 3) (SAP-3).

GN Name=GM2a;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;

RX MEDLINE=95229165; PubMed=7713516;

RA Yamanaka S., Johnson O.N., Lyu M.S., Kozak C.A., Proia R.L.;

RT "The mouse gene encoding the GM2 activator protein (Gm2a): cDNA

RT sequence, expression, and chromosome mapping.";

RL Genomics 24:601-604(1994).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=93371367; PubMed=7689829;

RA Bellachio G., Stirling J.L., Orlacchio A., Beccari T.;

RT "Cloning and sequence analysis of a cDNA clone coding for the mouse

RT GM2 activator protein.";

RL Biochem. J. 294:227-230(1993).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6 X CBA;

RX MEDLINE=97224573; PubMed=9060405;

RA Bertoni C., Appolloni M.G., Stirling J.L., Li S.C., Li Y.T.,

RA Orlacchio A., Beccari T.;

RT "Structural organization and expression of the gene for the mouse GM2

RT activator protein.";

RL Mamm. Genome 8:90-93(1997).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Mammary gland;

RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loguillano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Faney J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: Binds gangliosides and stimulates ganglioside GM2

CC degradation. It stimulates only the breakdown of ganglioside GM2

CC and glycolipid GA2 by beta-hexosaminidase A. It extracts single

CC GM2 molecules from membranes and presents them in soluble form to

CC beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and

CC conversion to GM3.

CC -!- SUBCELLULAR LOCATION: Lysosomal.

CC -!- TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney and

CC testis.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U09816; AAA21543.1; -

DR EMBL; L19526; AAA61929.1; -

DR EMBL; U34359; AAB06275.1; ALT SEQ.

DR EMBL; U34356; AAB06275.1; JOINED.

DR EMBL; U34357; AAB06275.1; JOINED.

DR EMBL; U34358; AAB06275.1; JOINED.

DR EMBL; BC004651; AAH04651.1; -

DR HSSP; P17900; 1G13.

DR MGD; MGI:95762; Gm2a.

KW Glycoprotein; Lysosome; Signal; Sphingolipid metabolism.

FT SIGNAL 1 31 By similarity.

FT CHAIN 32 193 Ganglioside GM2 activator.

FT DISULFID 39 183 By similarity.

FT DISULFID 99 106 By similarity.

FT DISULFID 112 138 By similarity.

FT DISULFID 125 136 By similarity.

FT CARBOHYD 151 151 N-linked (GLCNac...) (Potential).

FT CONFLICT 53 53 I -> T (in Ref. 1).

SQ SEQUENCE 193 AA; 20824 MW; 59CC4ABE56FA1FC7 CRC64;

Query Match 73.3%; Score 746.5; DB 1; Length 193;

Best Local Similarity 71.8%; Pred. No. 9.8e-60;

Matches 135; Conservative 24; Mismatches 26; Indels 3; Gaps 1;

QY 8 PLLIALGLLLA---TPQAHLKKPSQLSFSWDCDEGKOPAVIRSLTLEPPPIVVGNV 64

DB 5 PLLLLGLLLAGSVAPARLVKRLSGLGFSWDCDEGKOPAVIRSLTLEPPPIVVGVD 64

QY 65 TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMLPTGEP 124

DB 65 VVSLEKTSVPLTAPQKVELTVEKEVAGFWKIPCEQLGSCSYENICDLIDEYIPPGS 124

QY 125 CPEPLRTYGLPCHCPFKEGTYSLPKSEFVVPDLELPWSLTTNNYIESVLSSGKRLGCI 184

DB 125 CPEPLHTYGLPCHCPFKEGTYSLPNTSNTVPDLELPWSLSTNGYRIQSLSSGKRLGCI 184

QY 185 KIAASLKG 192

DB 185 KIAASLKG 192

RESULT 8

Q6Q7X5 PRELIMINARY; PRT; 151 AA.

AC Q6Q7X5; TREMBLrel. 27, Created)

DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)

DE GM2 activator protein (Fragment).

GN Name=GM2a;

OS Felis silvestris catus (Cat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.

OX NCBI_TaxID=9685;

RN [1]

RP SEQUENCE FROM N.A.

RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,

RA Dodson A.N., Gentry A.S., Griffin B., Rolfsma M.D., Baker H.J.;

RL Submitted (FEB-2004) to the EMBL/GenBank/DBS databases.

DR EMBL; AY563653; AAS64350.1; -

DR InterPro; IPR003172; El_DerP2_DerF2.

DR SMART; SM00737; ML; 1

FT NON TER 1

SQ SEQUENCE 151 AA; 16310 MW; E23EF16CE5844A5F CRC64;

Query Match 59.1%; Score 602; DB 2; Length 151;

Best Local Similarity 68.7%; Pred. No. 1e-46;

Matches 103; Conservative 24; Mismatches 23; Indels 0; Gaps 0;

QY 43 KDPVIRSLTLEPPPIVVGNVTLSSVSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDY 102

DR EMBL; X61034; CAA43407.1; -;

RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,

```
RA Dodson A.N., Gentry A.S., Griffin B., Rolasa M.D., Baker H.J.;  
RT "Mutation of the GM2 activator protein in a feline model of GM2  
RT gangliosidosis";  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY553654; AAS64351.1; --  
FT NON_TER 1  
SQ SEQUENCE 146 AA; 15820 MW; 754E06C0CFCFA74E3 CRC64;  
  
Query Match 51.0%; Score 519; DB 2; Length 146;  
Best Local Similarity 66.9%; Pred. No. 3.4e-39;  
Matches 87; Conservative 21; Mismatches 22; Indels 0; Gaps 0;  
  
QY 43 KDPVIRSLTLEPDPPIVPGNVTLSVVGSTVPLSSPLKVDLVLEKAVGLWIKIPCTDY 102  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
1 KDPVIRSLTLEPDPPIVPGNVTLSVVGSTVPLSSPLKVDLVLEKAVGLWIKIPCTDY 60  
  
QY 103 IGSCTFEHPCDVLDMILPTGEPCEPLRTYGLPCHCPFKEGTYSLPKSFVVPDLELPSW 162  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
61 IGSCTFEHPCDVLDMILPTGEPCEPLRTYGLPCHCPFKEGTYSLPKSFVVPDLELPSW 120  
  
QY 163 LTGNRYIES 172  
Db ||:|||||:  
121 LSSGHYRIKT 130  
  
RESULT 13  
Q6GLN6 PRELIMINARY; PRT; 197 AA.  
ID Q6GLN6  
AC Q6GLN6;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Hypothetical protein.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Eye;  
RX MEDLINE=2338257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grumwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywicki M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Eye;  
RX MEDLINE=22341132; PubMed=12454917;  
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
RA Richardson P.;  
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
RT initiative";  
RL Dev. Dyn. 225:384-391 (2002).  
RN [3]  
RP SEQUENCE FROM N.A.
```

```
RC TISSUE=Eye;  
RA Klein S., Strausberg R.;  
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC074424; AAH74424.1; --  
DR InterPro: IPR003172; EI_DerP2_DerP2.  
DR SMART; SM00737; ML; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 197 AA; 21401 MW; A0909CAD82FD66EB CRC64;  
  
Query Match 50.3%; Score 512.5; DB 2; Length 197;  
Best Local Similarity 51.4%; Pred. No. 1.9e-38;  
Matches 90; Conservative 34; Mismatches 50; Indels 1; Gaps 1;  
  
QY 17 LATPAQAHKKPSQSLSSFSWDCDEKQDPAVIRSLTLEPDPPIVPGNVTLSVVGSTVPL 76  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
18 LASEAASASFLS--VNGFWSNCDGESLPGKIKLSVNPINIPGDLTVSVLETQVPL 76  
  
QY 77 SSPLKVDLVLEKAVGLWIKIPCTDYIGSCTFEHPCDVLDMILPTGEPCEPLRTYGLPC 136  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
77 TSPVKVIITAEKELLGEMMKVPCLDNIGSCTYDNACELIDTIFPPGQQCPPEPLRTYGLPC 136  
  
QY 137 HCPFKEGTYSLPKSFVVPDLELPSWLTGNRYIESVSSSGKRLGCIKIAASLK 191  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
137 HCPFKEGVYSLPDTTILPNDLPVSLANGNRIITGILMADNKEIGCGKFTSLE 191  
  
RESULT 14  
Q75R48 PRELIMINARY; PRT; 217 AA.  
ID Q75R48  
AC Q75R48;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE YGSC-1.  
GN Name=YGSC-1;  
OS Lymnaea stagnalis (Great pond snail).  
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;  
OC Lymnaeidae; Lymnaeidae; Lymnaea.  
OX NCBI_TaxID=6523;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14986137;  
RA Harada Y., Hosoiri Y., Kuroda R.;  
RT "Isolation and evaluation of dextral-specific and dextral-enriched  
RT cDNA clones as candidates for the handedness-determining gene in a  
RT freshwater gastropod, Lymnaea stagnalis";  
RL Dev. Genes Evol. 214:159-169 (2004).  
DR EMBL; AB159153; BAD16601.1; --  
DR InterPro: IPR003172; EI_DerP2_DerP2.  
DR SMART; SM00737; ML; 1.  
SQ SEQUENCE 217 AA; 23436 MW; DD626A73362E38AB CRC64;  
  
Query Match 25.7%; Score 262; DB 2; Length 217;  
Best Local Similarity 37.7%; Pred. No. 1.2e-15;  
Matches 60; Conservative 20; Mismatches 69; Indels 10; Gaps 4;  
  
QY 33 SFSWDC-DEGKDPVIRSLTLEPDPPIVPGNVTLSVVGSTVPLSSPLKVDLVLEKVA 91  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
67 SFSFKNCDPDNEILVPSNFNLEPDPPIRAPGNITVSGNLEIKSGSPLVASVVMKKVL 126  
  
QY 92 GLWIKIPCTDYIGSCTFEHPCDVLDMILPTGEPCEPLRTYGLPCHCPFKEGTYSLPKSE 151  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
127 GIWIKIPCYHGVGCTYSDACTLL-----TSPDCPTVLTIKGLPCQCPFPAGTFNFPED 181  
  
QY 152 FVVPDLELPSWLTGNRYIESVSSSGKRLGCIKIAASL 190  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
182 IVIPK-ALP---VSGEIFHLKTSYEGSLVTCVDLQFEL 216  
  
RESULT 15  
BAD16601 PRELIMINARY; PRT; 217 AA.  
ID BAD16601  
AC BAD16601;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 07:25:04 ; Search time 31.2046 Seconds
(without alignments)
2218.735 Million cell updates/sec

Title: US-10-030-937-9
Perfect score: 193
Sequence: 1 MQSLMQAPLLIALGLLLATP.....LSSSGKRLGCIKIAASLKGI 193

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq 23Sep04:*

- 1: geneseqp1980s.*
- 2: geneseqp1980s.*
- 3: geneseqp2000s.*
- 4: geneseqp2001s.*
- 5: geneseqp2002s.*
- 6: geneseqp2003as.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193	100.0	193	4	AAB31897 Amino aci
2	112	58.0	178	4	AAB31898 Amino aci
3	112	58.0	178	5	ABG31346 Non-glyco
4	112	58.0	189	4	AAB31900 Amino aci
5	112	58.0	193	4	AAB31901 Amino aci
6	112	58.0	193	4	AAB31902 Amino aci
7	112	58.0	193	4	AAB31904 Amino aci
8	112	58.0	193	4	AAB31896 Amino aci
9	112	58.0	193	4	AAB31928 Amino aci
10	112	58.0	193	4	AAB31903 Amino aci
11	112	58.0	193	4	ABG00720 Novel hum
12	112	58.0	193	5	ABG31345 Human GM2
13	112	58.0	193	5	ABP65212 Hypoxia-x
14	112	58.0	193	7	ADN95858 Human BEC
15	112	58.0	193	8	ADN03620 Antipsori
16	112	58.0	193	8	ADQ17712 Human sof
17	83	43.0	200	4	AAB31899 Amino aci
18	75	38.9	76	4	ABG00717 Novel hum
19	61	31.6	61	4	AAM15082 Peptide #
20	61	31.6	61	4	ABB34073 Peptide #
21	61	31.6	61	4	AAM27530 Peptide #
22	61	31.6	61	4	ABB28897 Peptide #
23	61	31.6	61	4	ABB19511 Protein #
24	61	31.6	61	4	AAM67236 Human bon
25	61	31.6	61	4	AAM54855 Human bra

26	61	31.6	61	4	ABG48902	Abg48902 Human liv
27	61	31.6	61	4	AAM02818	Aam02818 Peptide #
28	61	31.6	61	5	ABG36887	Abg36887 Human pep
29	18	9.3	18	4	AAB31920	Aab31920 Amino aci
30	18	9.3	199	2	AAW10656	Aaw10656 Rat GM2 a
31	16	8.3	16	4	AAB31927	Aab31927 Amino aci
32	16	8.3	16	4	AAB31926	Aab31926 Amino aci
33	16	8.3	131	4	ABG00719	Abg00719 Novel hum
34	15	7.8	15	4	AAB31918	Aab31918 Amino aci
35	13	6.7	13	4	AAB31919	Aab31919 Amino aci
36	11	5.7	11	4	AAU25350	Aau25350 Schizophr
37	11	5.7	11	4	AAU15694	Aau15694 Schizophr
38	11	5.7	11	8	ADO78961	Ado78961 Schizophr
39	9	4.7	262	2	AAW55111	Aaw55111 Streptoco
40	9	4.7	262	5	ABP54605	Abp54605 S. pneumo
41	9	4.7	262	7	ADC45179	Adc45179 S. pneumo
42	9	4.7	270	2	AAW61241	Aaw61241 Streptoco
43	9	4.7	270	5	ABP54659	Abp54659 S. pneumo
44	9	4.7	270	7	ADC45287	Adc45287 S. pneumo
45	9	4.7	291	6	ABU02596	Abu02596 S. pneumo

ALIGNMENTS

RESULT 1
AAB31897
ID AAB31897 standard; protein; 193 AA.

XX AAB31897;

DT 15-MAY-2001 (first entry)

DE Amino acid sequence of a mutant ganglioside GM2 activator protein.

KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

OS Homo sapiens.

PN WO200105422-A2.

PD 25-JAN-2001.

PF 17-JUL-2000; 2000WO-FR002057.

PR 15-JUL-1999; 99FR-00009372.

PA (INMR) BIOMERIEUX STELHYS.

PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.

DR Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.

PS Claim 25; Page 159-160; 209pp; French.

XX The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and

CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;
 Query Match 100.0%; Score 193; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 2.5e-187;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIALGILLATPAQAHLKPSQSFNDNCFEGKDPVIRSLTLEPPIV 50
 DB 1 MQSLMQAPLLIALGILLATPAQAHLKPSQSFNDNCFEGKDPVIRSLTLEPPIV 60
 QY 61 PGNVTLVVVGSTSVPLSPKVDLVLEKEVAGLWIKIPCTDYIGSTCFEHCDFVLDMLIP 120
 DB 61 PGNVTLVVVGSTSVPLSPKVDLVLEKEVAGLWIKIPCTDYIGSTCFEHCDFVLDMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFAVPDLELPWLTGNYRIEIVLSSGKR 180
 DB 121 TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFAVPDLELPWLTGNYRIEIVLSSGKR 180
 QY 181 LGCIIKAASLKG 193
 DB 181 LGCIIKAASLKG 193

RESULT 2
 AAB31898
 ID AAB31898 standard; protein; 178 AA.
 XX
 AC
 XX
 DT 15-MAY-2001 (first entry)
 DE Amino acid sequence of a human protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHVS.
 XX
 PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 WPI; 2001-159475/16.
 XX
 DR Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 1; Page 160; 209pp; French.
 XX
 CC The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a

CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 178 AA;
 Query Match 58.0%; Score 112; DB 4; Length 178;
 Best Local Similarity 100.0%; Pred. No. 4.4e-105;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 EKGDPVIRSLTLEPPIVPGNVTLVVVGSTSVPLSPKVDLVLEKEVAGLWIKIPCT 100
 DB 26 EKGDPVIRSLTLEPPIVPGNVTLVVVGSTSVPLSPKVDLVLEKEVAGLWIKIPCT 85
 QY 101 DYIGSTCFEHCDFVLDMLIPTGCPPEPLRTYGLPCHCPFKEGTYSLPKSEF 152
 DB 86 DYIGSTCFEHCDFVLDMLIPTGCPPEPLRTYGLPCHCPFKEGTYSLPKSEF 137
 RESULT 3
 ABG31346
 ID ABG31346 standard; protein; 178 AA.
 XX
 AC ABG31346;
 XX
 DT 15-NOV-2002 (first entry)
 DE Non-glycosylated human GM2 activator protein containing His6-tag.
 XX
 KW Human; GM2 activator protein; ganglioside; platelet activating factor;
 KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
 KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
 KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
 KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
 KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
 KW antiinflammatory; antiasthmatic; antiarthritic; antirheumatic;
 KW antiallergic; hepatotropic; nephrotropic; immunosuppressive;
 KW tranquiliser.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PH Key Location/Qualifiers
 FT Binding-site 5..10
 FT /label= His6 tag
 FT /note= "Nickel chelating region used for purification of
 FT the protein"
 FT Cleavage-site 11..16
 FT /label= Factor_X_cleavage_site
 XX
 PN US6423680-B1.
 XX
 PD 23-JUL-2002.
 XX
 PF 30-OCT-1998; 98US-00183841.
 XX
 PR 30-OCT-1998; 98US-00183841.
 XX
 PA (HSCR-) HSC RES & DEV LP.
 XX
 PI Rigat B, Reynaud D, Mahuran D;
 XX
 WPI; 2002-664636/71.
 DR Composition useful for treating inflammatory conditions e.g. asthma
 PT comprises GM2 activator protein or GM2 activator peptide in combination
 PT with a carrier.

XX
PS Example 1; Fig 2; 11pp; English.
XX
XX The present invention relates to a composition comprising GM2
CC (ganglioside) activator protein or a GM2 activator peptide derived from
CC the GM2 activator protein in combination with a carrier. The composition
CC comprises the protein or peptide in an amount of 1-100 mg. The GM2
CC activator protein is capable of inhibiting platelet activating factor
CC (PAF). The composition of the invention is useful for treating
CC inflammatory disorders e.g. inflammatory bowel disease, asthma,
CC autoimmune disease (such as lupus), hypersensitivity infection,
CC rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
CC gout and tissue-specific conditions (e.g. glomerulonephritis and
CC hepatitis). The composition is capable of inhibiting platelet activating
CC factor, is non-toxic, is efficacious and presents less severe side
CC effects, including redness, swelling, pain and polymorphonuclear
CC leukocyte accumulation at the inflammatory site and other associated
CC cellular responses. The present sequence represents a non-glycosylated
CC human GM2 activator protein prepared using a His6-tag bacterial
CC expression system
XX
SQ Sequence 178 AA;
Query Match 58.0%; Score 112; DB 5; Length 178;
Best Local Similarity 100.0%; Pred. No. 4.4e-105;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 EGKDPVAVIRSLTLEPDPPIVVGNTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100
DB 26 EGKDPVAVIRSLTLEPDPPIVVGNTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 85
QY 101 DYIGSCTFEHCVDLMDLIPTGEPCEPLRTYGLPCHCPKFGTYSLPKSEF 152
DB 86 DYIGSCTFEHCVDLMDLIPTGEPCEPLRTYGLPCHCPKFGTYSLPKSEF 137
RESULT 4
AAB31900
ID AAB31900 standard; protein; 189 AA.
XX
AC AAB31900;
XX
DT 15-MAY-2001 (first entry)
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX W0200105422-A2.
XX 25-JAN-2001.
XX 17-JUL-2000; 2000WO-FR002057.
XX 15-JUL-1999; 99FR-00009372.
XX (INMR) BIOMERIEUX STELHYS.
XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand.
XX Claim 1; Page 161-162; 209pp; French.

XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 189 AA;
Query Match 58.0%; Score 112; DB 4; Length 189;
Best Local Similarity 100.0%; Pred. No. 4.6e-105;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 EGKDPVAVIRSLTLEPDPPIVVGNTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100
DB 37 EGKDPVAVIRSLTLEPDPPIVVGNTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 96
QY 101 DYIGSCTFEHCVDLMDLIPTGEPCEPLRTYGLPCHCPKFGTYSLPKSEF 152
DB 97 DYIGSCTFEHCVDLMDLIPTGEPCEPLRTYGLPCHCPKFGTYSLPKSEF 148
RESULT 5
AAB31901
ID AAB31901 standard; protein; 193 AA.
XX
AC AAB31901;
XX
DT 15-MAY-2001 (first entry)
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX W0200105422-A2.
XX 25-JAN-2001.
XX 17-JUL-2000; 2000WO-FR002057.
XX 15-JUL-1999; 99FR-00009372.
XX (INMR) BIOMERIEUX STELHYS.
XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand.
XX Claim 1; Page 162-163; 209pp; French.
XX The present sequence represents a human protein, which is used in the
XX method of the invention. The specification describes a method which uses
XX at least one polypeptide or polynucleotide sequence belonging to the

CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;

Query Match 58.0%; Score 112; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 4.7e-105;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 ECKDPAVIRSLTEPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100
DB 41 ECKDPAVIRSLTEPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100

QY 101 DYIGSTFHFCDVLDMLIPTGCPPEPLRTYGLPCHCFKGTSLPKSEF 152
DB 101 DYIGSTFHFCDVLDMLIPTGCPPEPLRTYGLPCHCFKGTSLPKSEF 152

RESULT 6
AAB31902
ID AAB31902 standard; protein; 193 AA.
XX
AC AAB31902;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
FN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR) BIOMERIEUX STELHYS.
XX
PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
DR Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 163; 209pp; French.
XX

The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The

CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;

Query Match 58.0%; Score 112; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 4.7e-105;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 ECKDPAVIRSLTEPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100
DB 41 ECKDPAVIRSLTEPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100

QY 101 DYIGSTFHFCDVLDMLIPTGCPPEPLRTYGLPCHCFKGTSLPKSEF 152
DB 101 DYIGSTFHFCDVLDMLIPTGCPPEPLRTYGLPCHCFKGTSLPKSEF 152

RESULT 7
AAB31904
ID AAB31904 standard; protein; 193 AA.
XX
AC AAB31904;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
FN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR) BIOMERIEUX STELHYS.
XX
PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
DR Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 164-165; 209pp; French.
XX

The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid

CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;
 Query Match 58.0%; Score 112; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 4.7e-105;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 EGKDPVAVIRSLTLEPPIVPGNVTLVSVGTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100
 DB 41 EGKDPVAVIRSLTLEPPIVPGNVTLVSVGTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100
 QY 101 DYIGSTCTFEHFCVDLMDLPTGCPPEPLRTYGLPCHCPKCTYSLPKSEF 152
 DB 101 DYIGSTCTFEHFCVDLMDLPTGCPPEPLRTYGLPCHCPKCTYSLPKSEF 152
 RESULT 8
 AAB31896
 ID AAB31896 standard; protein; 193 AA.
 XX
 AC AAB31896;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of the human ganglioside GM2 activator protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.
 XX
 DR N-PSDB; AAF54698.
 XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 23; Page 158-159; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly

CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;
 Query Match 58.0%; Score 112; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 4.7e-105;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 EGKDPVAVIRSLTLEPPIVPGNVTLVSVGTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100
 DB 41 EGKDPVAVIRSLTLEPPIVPGNVTLVSVGTSVPLSSPLKVDLVLEKEVAGLWIKIPCT 100
 QY 101 DYIGSTCTFEHFCVDLMDLPTGCPPEPLRTYGLPCHCPKCTYSLPKSEF 152
 DB 101 DYIGSTCTFEHFCVDLMDLPTGCPPEPLRTYGLPCHCPKCTYSLPKSEF 152
 RESULT 9
 AAB31928
 ID AAB31928 standard; protein; 193 AA.
 XX
 AC AAB31928;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of the human ganglioside GM2 activator protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.
 XX
 DR Detecting, preventing and treating degenerative, neurological and
 DR autoimmune diseases, particularly multiple sclerosis, using specified
 DR polypeptides or related nucleic acid or ligand.
 XX
 PS Disclosure; Fig 1; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 58.0%; Score 112; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 4.7e-105; Indels 0; Gaps 0;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 EKGDPVIRSLTLEPPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGIWKIPCT 100
 DB 41 EKGDPVIRSLTLEPPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGIWKIPCT 100

QY 101 DYIGSCTFHFCDVLDMLIPTGCPCEPLRTYGLPCHCFKCTYSLPKSEF 152
 DB 101 DYIGSCTFHFCDVLDMLIPTGCPCEPLRTYGLPCHCFKCTYSLPKSEF 152

RESULT 10
 AAB31903
 ID AAB31903 standard; protein; 193 AA.
 AC AAB31903;
 XX
 DT 15-MAY-2001 (first entry)
 DE Amino acid sequence of a human protein.
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR02057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INNR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.
 XX
 PS Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 1; Page 164; 209pp; French.
 XX
 CC The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 58.0%; Score 112; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 4.7e-105; Indels 0; Gaps 0;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 EKGDPVIRSLTLEPPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGIWKIPCT 100
 DB 41 EKGDPVIRSLTLEPPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGIWKIPCT 100

QY 101 DYIGSCTFHFCDVLDMLIPTGCPCEPLRTYGLPCHCFKCTYSLPKSEF 152
 DB 101 DYIGSCTFHFCDVLDMLIPTGCPCEPLRTYGLPCHCFKCTYSLPKSEF 152

RESULT 11
 ABG00720
 ID ABG00720 standard; protein; 193 AA.
 XX
 AC ABG00720;
 XX
 DT 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #711.
 DE
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 XX
 PS N-PSDB; AAS64907.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 31079; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 58.0%; Score 112; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 4.7e-105;

Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 EKGDAVIRSLTLEPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGIWKIPCT 100
 DB 41 EKGDAVIRSLTLEPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGIWKIPCT 100
 QY 101 DYIGSCTFEHFCVDLMDLPTGCPPEPLRTYGLPCHCFKSGTYSLPKSEF 152
 DB 101 DYIGSCTFEHFCVDLMDLPTGCPPEPLRTYGLPCHCFKSGTYSLPKSEF 152

RESULT 12
 ABG31345
 ID ABG31345 standard; protein; 193 AA.
 XX
 AC ABG31345;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Human GM2 activator protein.
 XX
 KW Human; GM2 activator protein; ganglioside; platelet activating factor;
 KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
 KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
 KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
 KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
 KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
 KW antiinflammatory; antiasthmatic; antiarthritic; antirheumatic;
 KW antiallergic; hepatotropic; nephrotropic; immunosuppressive;
 KW tranquiliser.

OS Homo sapiens.
 XX
 PN US6423680-B1.
 XX
 PD 23-JUL-2002.
 XX
 PF 30-OCT-1998; 98US-00183841.
 XX
 PR 30-OCT-1998; 98US-00183841.
 XX
 PA (HSCR-) HSC RES & DEV LP.
 XX
 XX Rigat B, Reynaud D, Mahuran D;
 XX WPI; 2002-664636/71.
 XX
 PT Composition useful for treating inflammatory conditions e.g. asthma
 PT comprises GM2 activator protein or GM2 activator peptide in combination
 PT with a carrier.
 XX
 PS Claim 3; Fig 1; 11pp; English.

XX
 CC The present invention relates to a composition comprising GM2
 CC (ganglioside) activator protein or a GM2 activator peptide derived from
 CC the GM2 activator protein in combination with a carrier. The composition
 CC comprises the protein or peptide in an amount of 1-100 mg. The GM2
 CC activator protein is capable of inhibiting platelet activating factor
 CC (PAF). The composition of the invention is useful for treating
 CC inflammatory disorders e.g. inflammatory bowel disease, asthma,
 CC autoimmune disease (such as lupus), hypersensitivity infection,
 CC rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
 CC gout and tissue-specific conditions (e.g. glomerulonephritis and
 CC hepatitis). The composition is capable of inhibiting platelet activating
 CC factor, is non-toxic, is efficacious and presents less severe side
 CC effects, including redness, swelling, pain and polymorphonuclear
 CC leukocyte accumulation at the inflammatory site and other associated
 CC cellular responses. The present sequence represents human GM2 activator
 CC protein

XX Sequence 193 AA;

Query Match 58.0%; Score 112; DB 5; Length 193;

Best Local Similarity 100.0%; Pred. No. 4.7e-105;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 EKGDAVIRSLTLEPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGIWKIPCT 100
 DB 41 EKGDAVIRSLTLEPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGIWKIPCT 100
 QY 101 DYIGSCTFEHFCVDLMDLPTGCPPEPLRTYGLPCHCFKSGTYSLPKSEF 152
 DB 101 DYIGSCTFEHFCVDLMDLPTGCPPEPLRTYGLPCHCFKSGTYSLPKSEF 152

RESULT 13
 ABP65212
 ID ABP65212 standard; protein; 193 AA.
 XX
 AC ABP65212;
 XX
 DT 12-NOV-2002 (first entry)
 XX
 DE Hypoxia-regulated protein #86.
 XX
 KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;
 KW antiinflammatory; vulnery; gynecological; ophthalmological; vaccine;
 KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
 KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;
 KW preclampsia; atherosclerosis; inflammatory condition; wound healing;
 KW inflammation; erythropoiesis; hair loss; human.

XX Homo sapiens.

XX WO200246465-A2.

XX 13-JUN-2002.

XX 10-DEC-2001; 2001WO-GB005458.

XX 08-DEC-2000; 2000GB-00030076.

XX 08-FEB-2001; 2001GB-00003156.

XX 25-OCT-2001; 2001GB-00025666.

XX (OXFO-) OXFORD BIOMEDICA UK LTD.

XX White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;
 XX Rayner WN;

XX WPI; 2002-627238/67.

XX Identifying a gene involved in disease for treating hypoxia-regulated
 PT conditions, comprises comparing the transcriptome/proteome of two cell
 PT types under different conditions and identifying a differentially
 PT regulated gene.

XX Claim 35; Page 397; 538pp; English.

XX The present invention relates to methods for identifying genes and
 CC proteins that are implicated in a specific disease or physiological
 CC condition. The method comprises comparing the transcriptome/proteome of a
 CC specialised cell type implicated in a disease or condition with that of a
 CC second specialised cell type, under two experimental conditions, and
 CC identifying a gene that is differentially regulated in the two
 CC specialised cell types under experimental conditions. ABV77873-ABV78116
 CC and ABP65061-ABP65257 were identified using the methods of the invention.
 CC The coding sequences and proteins are useful for treating a disease in a
 CC patient, for manufacture of a medicament for treating hypoxia-regulated
 CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
 CC biological response to hypoxia conditions, or hypoxic-associated
 CC pathology in a patient. The coding sequences and proteins are also useful
 CC for monitoring the therapeutic treatment of a disease or physiological
 CC condition, such as cancer, ischaemic conditions, reperfusion injury,
 CC retinopathy, neonatal stress, preclampsia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss

SQ		Sequence 193 AA;			
Query Match		58.0%; Score 112; DB 5; Length 193;			
Best Local Similarity		100.0%; Pred. No. 4.7e-105;			
Matches 112; Conservative		0; Mismatches 0; Indels 0; Gaps 0;			
QY	41	EGKDPVIRSLTLEPDPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT	100		
Db	41	EGKDPVIRSLTLEPDPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT	100		
QY	101	DYIGSCTFEHFCVDLMDLIPTGEPCEPLRTYGLPCHCPFKGTYSLPKSEF	152		
Db	101	DYIGSCTFEHFCVDLMDLIPTGEPCEPLRTYGLPCHCPFKGTYSLPKSEF	152		
RESULT 14		(first entry)			
ADN95858		Human BEC/LEC-related protein sequence SeqID782.			
ID	ADN95858	standard; protein; 193 AA.			
AC	ADN95858;				
XX	01-JUL-2004	(first entry)			
DE	growth; differentiation; blood endothelial cell; BEC;				
KW	lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;				
KW	lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;				
KW	vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;				
KW	inflammatory disease; cancer metastasis; lymphatic system; human.				
XX	Homo sapiens.				
OS	WO2003080640-A1.				
PN	02-OCT-2003.				
XX	07-MAR-2003; 2003WO-US006900.				
XX	07-MAR-2002; 2002US-0363019P.				
XX	(LUDW-) LUDWIG INST CANCER RES.				
PA	(LICN) LICENTIA LTD.				
XX	Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;				
PI	WPI; 2003-876899/81.				
DR	N-PSDB; ADN95859.				
XX	Example 1; SEQ ID NO 782; 176pp; English.				
PS	This invention relates to a method of differentially modulating the				
CC	growth or differentiation of blood endothelial cells (BEC) or lymphatic				
CC	endothelial cells (LEC) comprises contacting endothelial cells with a				
CC	composition comprising an agent that differentially modulates blood or				
CC	lymphatic endothelial cells. Treating hereditary lymphoedema comprises				
CC	identifying a human subject with lymphoedema and with a mutation in at				
CC	least one allele of a gene encoding a LEC protein, where the mutation				
CC	correlates with lymphoedema in human subjects, and with the proviso that				
CC	the LEC protein is not VEGFR-3; and administering to the subject a				
CC	composition comprising a lymphatic growth agent selected from VEGF-C or				
CC	VEGF-D polypeptides and polynucleotides. The invention may be useful for				
CC	the development of compounds with an antiangiogenic, cytostatic,				
CC	vasotropic or antiinflammatory activity or for gene therapy. The method				
CC	is useful in modulating the growth or differentiation of blood				
CC	endothelial cells or lymphatic endothelial cells, in treating hereditary				
CC	lymphoedema, in screening for an endothelial cell disorder or				
CC	predisposition to the disorder or in monitoring the efficacy or toxicity				
CC	of a drug on endothelial cells. The agent is useful in manufacturing a				
CC	medicament for the differential modulation of blood vessel endothelial				
CC	cell or lymphatic vessel endothelial cell growth or differentiation. The				
CC	lymphatic growth agent may also be used in manufacturing a medicament for				
CC	the treatment of hereditary lymphoedema resulting from a mutation in a				
CC		LEC gene or of other diseases involving the lymphatic vessels, such as			
CC		various inflammatory diseases and cancer metastasis via the lymphatic			
CC		system. The present sequence is that of a human LEC/BEC differentially			
CC		expressed protein which is related to the method of the invention. Note:			
CC		This sequence does not appear in the specification but was obtained by			
CC		the indexer using the source data given in table 14 of the specification.			
XX		Sequence 193 AA;			
SQ		Query Match			
		58.0%; Score 112; DB 7; Length 193;			
		Best Local Similarity 100.0%; Pred. No. 4.7e-105;			
		Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	41	EGKDPVIRSLTLEPDPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT	100		
Db	41	EGKDPVIRSLTLEPDPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT	100		
QY	101	DYIGSCTFEHFCVDLMDLIPTGEPCEPLRTYGLPCHCPFKGTYSLPKSEF	152		
Db	101	DYIGSCTFEHFCVDLMDLIPTGEPCEPLRTYGLPCHCPFKGTYSLPKSEF	152		
RESULT 15		ADN03620			
ID	ADN03620	standard; protein; 193 AA.			
XX	ADN03620;				
XX	01-JUL-2004	(first entry)			
DE	Antipsoriatic protein sequence #7.				
XX	antipsoriatic; gene therapy; psoriasis; diagnosis.				
KW	Homo sapiens.				
OS	WO2004028479-A2.				
PN	08-APR-2004.				
XX	25-SEP-2003; 2003WO-US030907.				
XX	25-SEP-2002; 2002US-0414006P.				
XX	(GETH) GENENTECH INC.				
PI	Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;				
PI	Wu TD;				
XX	WPI; 2004-305105/28.				
DR	N-PSDB; ADN03619.				
XX	New PRO nucleic acid or polypeptide, useful for preparing a				
PT	pharmaceutical composition for diagnosing or treating psoriasis in a				
PT	mammal.				
PS	Claim 9; SEQ ID NO 14; 3069pp; English.				
XX	The invention relates to novel polynucleotide and polypeptides for				
CC	treating psoriasis or a sequence having at least 80% identity to the				
CC	above sequences. The nucleic acid is useful for preparing a composition				
CC	for diagnosing or treating psoriasis in a mammal. This sequence				
CC	corresponds to one of the polypeptides of the invention.				
XX		Sequence 193 AA;			
SQ		Query Match			
		58.0%; Score 112; DB 8; Length 193;			
		Best Local Similarity 100.0%; Pred. No. 4.7e-105;			
		Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	41	EGKDPVIRSLTLEPDPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT	100		
Db	41	EGKDPVIRSLTLEPDPPIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCT	100		

Qy 101 DYIGSCTFEHFCVLDMLIPTGEPCEPLRTYGLPCHCPFKEGTYSLPKSEF 152
Db 101 DYIGSCTFEHFCVLDMLIPTGEPCEPLRTYGLPCHCPFKEGTYSLPKSEF 152

Search completed: November 17, 2004, 09:12:38
Job time : 32.2046 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 10:19:31 ; Search time 7.70454 Seconds
(without alignments)
1661.276 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 193

Sequence: 1 MSLMQAPLLIALGLLATP.....LSSSGKRLGCIKIAASLKGI 193

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	58.0	178	4	US-09-183-841-2
2	112	58.0	193	4	US-09-183-841-1
3	9	4.7	262	3	US-08-961-083-98
4	9	4.7	262	4	US-09-536-784-98
5	9	4.7	270	3	US-08-961-083-206
6	9	4.7	270	4	US-09-536-784-206
7	9	4.7	291	4	US-09-583-110-4201
8	8	4.1	444	4	US-09-252-991A-20496
9	8	4.1	563	4	US-09-252-991A-23790
10	8	4.1	769	3	US-09-320-878-12
11	8	4.1	769	4	US-09-141-908-10
12	8	4.1	769	4	US-09-657-440-12
13	8	4.1	809	3	US-09-105-537-24
14	8	4.1	3782	3	US-09-105-537-4
15	7	3.6	85	4	US-09-270-767-32129
16	7	3.6	85	4	US-09-270-767-47346
17	7	3.6	96	3	US-08-936-165A-401
18	7	3.6	103	4	US-09-270-767-33345
19	7	3.6	103	4	US-09-270-767-48562
20	7	3.6	115	4	US-09-746-801A-47
21	7	3.6	153	3	US-09-199-637A-213
22	7	3.6	170	4	US-09-205-258-1011
23	7	3.6	267	4	US-09-489-039A-12889
24	7	3.6	269	4	US-09-543-681A-6475
25	7	3.6	285	1	US-08-149-809-24
26	7	3.6	287	4	US-09-540-336-2879
27	7	3.6	315	4	US-09-252-991A-31850

28	7	3.6	324	4	US-09-489-039A-7803
29	7	3.6	325	2	US-08-828-242-4
30	7	3.6	325	3	US-09-206-499-4
31	7	3.6	331	2	US-08-828-242-3
32	7	3.6	331	2	US-08-910-927B-5
33	7	3.6	331	2	US-09-206-499-3
34	7	3.6	331	3	US-09-270-270-5
35	7	3.6	331	4	US-09-961-403-11
36	7	3.6	355	4	US-09-580-929-5
37	7	3.6	364	4	US-09-205-258-1008
38	7	3.6	369	4	US-09-489-039A-8053
39	7	3.6	370	4	US-09-543-681A-4353
40	7	3.6	381	3	US-09-257-580-2
41	7	3.6	492	4	US-09-252-991A-32203
42	7	3.6	536	3	US-09-188-930-185
43	7	3.6	536	4	US-09-312-283C-185
44	7	3.6	590	4	US-09-312-283C-409
45	7	3.6	605	4	US-09-252-991A-24349

ALIGNMENTS

RESULT 1
US-09-183-841-2
; Sequence 2, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680e1 Inhibitor of Platelet Activating Factor
; FILE REFERENCE: van20010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: His tag at residues 1 to 17
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: amino acid
; OTHER INFORMATION: sequence of GM2 protein using His6 tag
US-09-183-841-2

Query Match	58.0%	Score 112;	DB 4;	Length 178;
Best Local Similarity	100.0%	Pred. No. 2.9e-102;		
Matches 112;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	41	EGKDPVIRSLTLEPDPPIVVPGNVTLVSVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCT 100		
Db	26	EGKDPVIRSLTLEPDPPIVVPGNVTLVSVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCT 85		
Qy	101	DYTGSCTFEHCVDLMDLIPTGCPPEPLRTYGLPCHCPKEGTYSLPKSEF 152		
Db	86	DYTGSCTFEHCVDLMDLIPTGCPPEPLRTYGLPCHCPKEGTYSLPKSEF 137		

RESULT 2
US-09-183-841-1
; Sequence 1, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680e1 Inhibitor of Platelet Activating Factor
; FILE REFERENCE: van20010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 193

```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (33)..(55)
; FEATURE:
; OTHER INFORMATION: residues 56-63 are included in a further precursor
; OTHER INFORMATION: form of the protein
US-09-183-841-1

Query Match          58.0%; Score 112; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 3.1e-102; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 0;

QY 41 EKKDPAVIRSLTEPDPVIVPGNVTLVVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPT 100
Db 41 EKKDPAVIRSLTEPDPVIVPGNVTLVVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPT 100

QY 101 DYIGSCTEHCVDLMDLIPTGCPPEPLRTYGLCHCPFKGTYSLPKSEF 152
Db 101 DYIGSCTEHCVDLMDLIPTGCPPEPLRTYGLCHCPFKGTYSLPKSEF 152

RESULT 3
US-08-961-083-98
; Sequence 98, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; QUERY MATCH 4.7%; Score 9; DB 3; Length 262;
; Best Local Similarity 100.0%; Pred. No. 0.87;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LSVVGSTSV 74
Db 155 LSVVGSTSV 163

US-08-961-083-98
```

```
RESULT 4
US-09-536-784-98
; Sequence 98, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 98:
US-09-536-784-98

Query Match          4.7%; Score 9; DB 4; Length 262;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LSVVGSTSV 74
Db 155 LSVVGSTSV 163

RESULT 5
US-08-961-083-206
; Sequence 206, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
```

```
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 206:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-206
```

```
Query Match 4.7%; Score 9; DB 3; Length 270;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 66 LSVVGSTSV 74
Db 163 LSVVGSTSV 171
```

```
RESULT 6
US-09-536-784-206
; Sequence 206, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 206:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 206:
US-09-536-784-206
Query Match 4.7%; Score 9; DB 4; Length 270;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 66 LSVVGSTSV 74
Db 163 LSVVGSTSV 171
RESULT 7
US-09-583-110-4201
; Sequence 4201, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4201
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4201
Query Match 4.7%; Score 9; DB 4; Length 291;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 66 LSVVGSTSV 74
Db 184 LSVVGSTSV 192
RESULT 8
US-09-252-991A-20496
; Sequence 20496, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20496
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (31)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-20496
```

Query Match 4.1%; Score 8; DB 4; Length 444;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 IALGLLLA 18
Db 162 IALGLLLA 169

RESULT 9
US-09-252-991A-23790
; Sequence 23790, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ASRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23790
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23790

Query Match 4.1%; Score 8; DB 4; Length 563;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 GLLLATPA 21
Db 60 GLLLATPA 67

RESULT 10
US-09-320-878-12
; Sequence 12, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-12

Query Match 4.1%; Score 8; DB 3; Length 769;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 GLLLATPA 21
Db 285 GLLLATPA 292

RESULT 11
US-09-141-908-10
; Sequence 10, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-141-908-10

Query Match 4.1%; Score 8; DB 4; Length 769;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 GLLLATPA 21
Db 285 GLLLATPA 292

RESULT 12
US-09-657-440-12
; Sequence 12, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR FILING DATE: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 769
; TYPE: PRT

; ORGANISM: Streptomyces venezuelae
US-09-657-440-12

Query Match 4.1%; Score 8; DB 4; Length 769;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GLLLATPA 21
Db 285 GLLLATPA 292

RESULT 13

US-09-105-537-24
; Sequence 24, Application US/09105537A

; Patent No. 6265202

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/105,537A

; CURRENT FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 24

; LENGTH: 809

; TYPE: PRT

; ORGANISM: Streptomyces venezuelae

US-09-105-537-24

Query Match 4.1%; Score 8; DB 3; Length 809;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GLLLATPA 21
Db 344 GLLLATPA 351

RESULT 14

US-09-105-537-4
; Sequence 4, Application US/09105537A

; Patent No. 6265202

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/105,537A

; CURRENT FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 4

; LENGTH: 3782

; TYPE: PRT

; ORGANISM: Streptomyces venezuelae

US-09-105-537-4

Query Match 4.1%; Score 8; DB 3; Length 3782;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GLLLATPA 21
Db 1409 GLLLATPA 1416

RESULT 15

US-09-270-767-32129
; Sequence 32129, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 32129

; LENGTH: 85

; TYPE: PRT

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-32129

Query Match 3.6%; Score 7; DB 4; Length 85;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 LEPLSWL 163
Db 53 LEPLSWL 59

Search completed: November 17, 2004, 10:53:16
Job time : 8.70454 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 10:45:47 ; Search time 25.07 Seconds

(without alignments)
2726.237 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 193

Sequence: 1 MSQLMQAPLLIALGILLATP.....LSSSGKRLGCKIAASLKGI 193

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1570615 seqs, 354127592 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	112	58.0	193	US-10-170-385-389
2	61	31.6	61	Sequence 389, App
3	11	5.7	9	Sequence 34809, A
4	11	5.7	11	Sequence 579, App
5	9	4.7	262	Sequence 579, App
6	9	4.7	9	Sequence 98, Appl
7	8	4.1	9	Sequence 206, App
8	8	4.1	8	Sequence 578, App
9	8	4.1	100	Sequence 578, App
10	8	4.1	110	Sequence 228216, App
11	8	4.1	170	Sequence 305353, App
12	8	4.1	323	Sequence 37, Appl
13	8	4.1	466	Sequence 5722, App
			16	Sequence 119990, App

14	8	4.1	574	16	US-10-324-967-36	Sequence 36, Appl
15	8	4.1	589	15	US-10-424-599-245422	Sequence 245422, App
16	8	4.1	769	10	US-09-793-708-12	Sequence 12, Appl
17	8	4.1	769	14	US-10-201-365-10	Sequence 10, Appl
18	8	4.1	769	14	US-10-160-539-12	Sequence 12, Appl
19	8	4.1	809	9	US-09-861-289-24	Sequence 24, Appl
20	8	4.1	809	9	US-09-860-846-24	Sequence 24, Appl
21	8	4.1	809	10	US-09-988-384B-24	Sequence 24, Appl
22	8	4.1	809	10	US-09-836-821-24	Sequence 24, Appl
23	8	4.1	809	14	US-10-271-889-24	Sequence 24, Appl
24	8	4.1	809	16	US-10-398-605-24	Sequence 24, Appl
25	8	4.1	1308	16	US-10-437-963-186215	Sequence 186215, App
26	8	4.1	3782	9	US-09-861-289-4	Sequence 4, Appl
27	8	4.1	3782	9	US-09-860-846-4	Sequence 4, Appl
28	8	4.1	3782	10	US-09-988-384B-4	Sequence 4, Appl
29	8	4.1	3782	10	US-09-836-821-4	Sequence 4, Appl
30	8	4.1	3782	14	US-10-271-889-47	Sequence 47, Appl
31	8	4.1	3782	16	US-10-398-605-4	Sequence 4, Appl
32	7	3.6	39	11	US-09-833-245-1222	Sequence 1222, App
33	7	3.6	39	11	US-09-833-245-1224	Sequence 1224, App
34	7	3.6	53	15	US-10-424-599-268608	Sequence 268608, App
35	7	3.6	54	17	US-10-425-115-220040	Sequence 220040, App
36	7	3.6	60	15	US-10-424-599-208921	Sequence 208921, App
37	7	3.6	60	15	US-10-424-599-241819	Sequence 241819, App
38	7	3.6	73	14	US-10-106-698-6961	Sequence 6961, App
39	7	3.6	83	15	US-10-425-114-61932	Sequence 61932, A
40	7	3.6	84	17	US-10-425-115-202828	Sequence 202828, App
41	7	3.6	87	14	US-10-029-386-31246	Sequence 31246, A
42	7	3.6	89	16	US-10-437-963-143338	Sequence 143338, App
43	7	3.6	92	16	US-10-437-963-152413	Sequence 152413, App
44	7	3.6	93	17	US-10-425-115-249956	Sequence 249956, App
45	7	3.6	96	9	US-09-939-980-401	Sequence 401, App

ALIGNMENTS

RESULT 1

US-10-170-385-389

; Sequence 389, Application US/10170385

; Publication No. US20030203372A1

; GENERAL INFORMATION:

; APPLICANT: Ward, Neil Raymond

; APPLICANT: Mundy, Christopher Robert

; APPLICANT: Kan, On

; APPLICANT: Harris, Robert Alan

; APPLICANT: White, Jonathan

; APPLICANT: Binley, Katie Mary

; APPLICANT: Rayner, William Nigel

; APPLICANT: Naylor, Stuart

; APPLICANT: Kingsman, Susan Mary

; APPLICANT: Krige, David

; TITLE OF INVENTION: ANALYSIS METHOD

; FILE REFERENCE: 532682000100

; CURRENT APPLICATION NUMBER: US/10/170,385

; CURRENT FILING DATE: 2002-06-12

; PRIOR APPLICATION NUMBER: PCT/GB02/01662

; PRIOR FILING DATE: 2002-04-08

; PRIOR APPLICATION NUMBER: PCT/GB01/05458

; PRIOR FILING DATE: 2001-12-10

; NUMBER OF SEQ ID NOS: 549

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 389

; LENGTH: 193

; TYPE: PRT

; ORGANISM: Homo Sapiens

US-10-170-385-389

Query Match 58.0%; Score 112; DB 14; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.7e-99;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 EGKDPVIRSLTEPDPVVPNGVNTLSVGVSTPLSKVDLVLEKEVAGLWIKIPCT 100

```

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99
; OTHER INFORMATION: EST HUMAN HIT: BE182886.1, EVALU9 9.00e-34
; OTHER INFORMATION: SWISSPROT HIT: P17900, EVALU9 1.00e-34
US-09-864-761-34809

```

[illegible]

```

RESULT 3
US-09-791-378-579
; Sequence 579, Application US/09791378
; Patent No. US20020142303A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF SCHIZOPHRENIA
; TITLE OF INVENTION: SCHIZOPHRENIA
; FILE REFERENCE: 9195-061-999
; CURRENT APPLICATION NUMBER: US/09/791,378
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 579
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-378-579

Query Match          5.7%; Score 11; DB 9; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      131 TYGLPCHCPEK 141
      |||||
Db      1 TYGLPCHCPEK 11

RESULT 4
US-09-791-377-579
; Sequence 579, Application US/09791377
; Publication No. US20040110938A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF SCHIZOPHRENIA
; TITLE OF INVENTION: SCHIZOPHRENIA
; FILE REFERENCE: 9195-060-999
; CURRENT APPLICATION NUMBER: US/09/791,377
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 579
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-377-579

Query Match          5.7%; Score 11; DB 11; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.003;

```

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 TYGLPCHCPEK 141
|||
Db 1 TYGLPCHCPEK 11

RESULT 5

```

US-09-765-272-98
; Sequence 98, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
;
; APPLICANT: Choi et. al.
;
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
;
; NUMBER OF SEQUENCES: 452
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSER: Human Genome Sciences, Inc.
;
; STREET: 9410 Key West Avenue
;
; CITY: Rockville
;
; STATE: Maryland
;
; COUNTRY: USA
;
; ZIP: 20850
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;
; COMPUTER: HP Vectra 486/33
;
; OPERATING SYSTEM: MSDOS version 6.2
;
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/09/765,272
;
; FILING DATE: 22-Jan-2001
;
; CLASSIFICATION: <Unknown>
;

```

RESULT 6

US-09-765-272-206
; Sequence 206, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland

—

RESOL 8
US-09-791-377-578
: Sequence 578, Application US/09791377

```
; Publication No. US20040110938A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 9195-060-999
; CURRENT APPLICATION NUMBER: US/09/791,377
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 578
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-377-578

Query Match      4.1%; Score 8; DB 11; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      89 EVAGLWIK 96
DB      1 EVAGLWIK 8

RESULT 9
US-10-424-599-228216
; Sequence 228216, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 228216
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_48107C.1.pap
US-10-424-599-228216

Query Match      4.1%; Score 8; DB 15; Length 100;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      71 STSVPLSS 78
DB      10 STSVPLSS 17

RESULT 10
US-10-425-115-305353
; Sequence 305353, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
```

```
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 305353
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_41551C.1.pap
US-10-425-115-305353

Query Match      4.1%; Score 8; DB 17; Length 110;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      48 IRSLTLEP 55
DB      19 IRSLTLEP 26

RESULT 11
US-09-804-014A-37
; Sequence 37, Application US/09804014A
; Publication No. US20030064489A1
; GENERAL INFORMATION:
; APPLICANT: Li, Li
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Vernet, Corine
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard
; APPLICANT: Spaderna, Steven
; APPLICANT: Majumder, Kumud
; TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-721 US
; CURRENT APPLICATION NUMBER: US/09/804,014A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/188,316
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/188,277
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/189,139
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/189,140
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/190,401
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/190,231
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-804-014A-37

Query Match      4.1%; Score 8; DB 10; Length 170;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      86 LEKEVAGL 93
DB      107 LEKEVAGL 114

RESULT 12
US-10-739-930-5722
; Sequence 5722, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
```

```
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 5722
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C126388_1.p
US-10-739-930-5722

Query Match      4.1%; Score 8; DB 17; Length 323;
Best Local Similarity 100.0%; Pred. No. 50;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      86 LEKEVAGL 93
Db      149 LEKEVAGL 156

RESULT 13
US-10-437-963-119990
; Sequence 119990, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 119990
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_23152C.1.p
US-10-437-963-119990

Query Match      4.1%; Score 8; DB 16; Length 466;
Best Local Similarity 100.0%; Pred. No. 69;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      86 LEKEVAGL 93
Db      337 LEKEVAGL 344

RESULT 14
US-10-324-967-36
; Sequence 36, Application US/10324967
; Publication No. US20040122212A1
; GENERAL INFORMATION:
; APPLICANT: Cosson, Pierre
; APPLICANT: Kohler, Thilo
; APPLICANT: Benghezal, Mohammed
; APPLICANT: Marchetti, Anna
; APPLICANT: van Delden, Christian
; TITLE OF INVENTION: VIRULENCE GENES, PROTEINS, AND THEIR USE
; FILE REFERENCE: 25421-502
; CURRENT APPLICATION NUMBER: US/10/324,967
; CURRENT FILING DATE: 2002-12-20
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 574

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-324-967-36

Query Match      4.1%; Score 8; DB 16; Length 574;
Best Local Similarity 100.0%; Pred. No. 84;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      11 IALGLLLA 18
Db      154 IALGLLLA 161

RESULT 15
US-10-424-599-245422
; Sequence 245422, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 245422
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_63647C.1.p
US-10-424-599-245422

Query Match      4.1%; Score 8; DB 15; Length 589;
Best Local Similarity 100.0%; Pred. No. 86;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      45 PAVIRSLT 52
Db      321 PAVIRSLT 328

Search completed: November 17, 2004, 12:23:48
Job time : 26.07 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 07:38:34 ; Search time 5.45839 Seconds
(without alignments)
3402.071 Million cell updates/sec

Title: US-10-030-937-9
Perfect score: 193
Sequence: 1 MQSLMQAPLLIALGILLATP.....LSSSGKRLGCIKIAASLKGI 193

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	58.0	193	2 I54178	ganglioside M2 act
2	83	43.0	162	2 S13195	ganglioside M2 act
3	83	43.0	200	2 S22411	ganglioside M2 act
4	18	9.3	193	2 S35613	ganglioside M2 act
5	9	4.7	291	2 G95243	hypothetical prote
6	9	4.7	312	2 D98108	hypothetical prote
7	8	4.1	20	2 S56005	lysosomal prote
8	8	4.1	163	2 B75402	hypothetical prote
9	8	4.1	300	2 S49337	hypothetical prote
10	8	4.1	357	2 B81396	fedF protein - Esc
11	8	4.1	389	2 A40809	probable aminotran
12	8	4.1	558	2 R82984	enamelin, 44K - bo
13	8	4.1	564	2 AH2321	hypothetical prote
14	8	4.1	574	2 H83117	hypothetical prote
15	7	3.6	96	2 T04223	probable ATP-bind
16	7	3.6	134	2 A87423	hypothetical prote
17	7	3.6	162	2 S17803	hypothetical prote
18	7	3.6	165	2 F72467	hypothetical prote
19	7	3.6	174	2 E84378	hypothetical prote
20	7	3.6	191	2 I69574	transposase - Esc
21	7	3.6	195	2 H83885	hypothetical prote
22	7	3.6	197	2 B69447	hypothetical prote
23	7	3.6	199	2 T08902	hypothetical prote
24	7	3.6	204	2 T34944	manganese-binding
25	7	3.6	210	2 B70740	probable lipoprote
26	7	3.6	218	2 E64004	hypothetical prote
27	7	3.6	224	2 E70386	hypothetical prote
28	7	3.6	226	2 F69125	flagellar motor pr
29	7	3.6	237	2 A41843	conserved hypothet
					protein xpsM - Xan

ALIGNMENTS

RESULT 1

I54178
ganglioside M2 activator protein precursor - human
C:Species: Homo sapiens (man)
C:Date: 24-May-1996 #sequence revision 24-May-1996 #text_change 09-Jul-2004
C:Accession: I54178; JQ1037; S05036; S22410; S17107
R:Xie, B.; Kennedy, J.L.; McInnes, B.; Auger, D.; Mahuran, D.
Genomics 14, 796-798, 1992
A:Title: Identification of a processed pseudogene related to the functional gene encoding man chromosome 5.
A:Reference number: I54178; MUID:93052421; PMID:1427911
A:Accession: I54178
A>Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-193 <RES>
A:Cross-references: UNIPROT:PI7900; GB:I01439; NID:g183358; PIDN:AAA52767.1; PID:g183359
R:Xie, B.; McInnes, B.; Neote, K.; Lamhounwah, A.M.; Mahuran, D.
Biochem. Biophys. Res. Commun. 177, 1217-1223, 1991
A:Title: Isolation and expression of a full-length cDNA encoding the human GM2 activator
A:Reference number: JQ1037; MUID:91282768; PMID:2059210
A:Accession: JQ1037
A:Molecule type: mRNA
A:Residues: 1-18, 'A', 20-193 <XIE>
A:Cross-references: GB:M76477; NID:g183356; PIDN:AAA35907.1; PID:g183357
A:Experimental source: HeLa cell
A:Note: 19-Thr and 69-Met were also found
R:Schroeder, M.; Klima, H.; Nakano, T.; Kwon, H.; Quintern, L.E.; Gaertner, S.; Suzuki, I.
FEBS Lett. 251, 197-200, 1989
A:Title: Isolation of a cDNA encoding the human G(M2) activator protein.
A:Reference number: S05036; MUID:89325664; PMID:2753159
A:Accession: S05036
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 15-193 <SCH>
A:Cross-references: GB:X16087; NID:g31852; PIDN:CAA34215.1; PID:g31853
R:Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A:Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A:Reference number: S22410; MUID:92207171; PMID:1554364
A:Accession: S22410
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 43-142, 'VST' <NAG>
A:Cross-references: EMBL:X61094; NID:g31854; PIDN:CAA43407.1; PID:g31855
C:Comment: This protein transports GM2 ganglioside from the lysosomal membrane to hexosan
C:Genetics:
A:Gene: GDB:GM2A
A:Cross-references: GDB:120000; OMIM:272750
A:Map position: 5q32-5q33
F:1-23/Domain: signal sequence #status predicted <SIG>
F:32-193/Product: GM2 ganglioside activator protein #status predicted <GM2>
F:32-193/Product: ganglioside M2 activator #status predicted <MAT>

30 7 3.6 239 2 A56808
31 7 3.6 245 2 AF0271
32 7 3.6 269 1 A30768
33 7 3.6 272 2 AI2005
34 7 3.6 273 2 E95268
35 7 3.6 273 2 S01167
36 7 3.6 285 2 C25242
37 7 3.6 285 2 A25561
38 7 3.6 285 2 H95969
39 7 3.6 293 2 A97396
40 7 3.6 293 2 AB2614
41 7 3.6 297 2 B83681
42 7 3.6 297 2 C83143
43 7 3.6 305 2 B75164
44 7 3.6 310 2 T41640
45 7 3.6 318 2 H36791

halorhodopsin - Ha
orotidine 5'-phosp
tryptophan synthas
delta-9 desaturase
probable ABC trans
cytochrome b561 -
tropomyosin, exon
tropomyosin II, mu
hypothetical prote
hypothetical prote
hypothetical prote
ABC transporter (p
hypothetical prote
hypothetical prote
probable checkpoin
hypothetical prote

```
Query Match      58.0%; Score 112; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 9.5e-107;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 EGKQPAVRSITLPPDPRIWVGNTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWKIPCT 100
    |||||
Db 41 EGKQPAVRSITLPPDPRIWVGNTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWKIPCT 100
    |||||

QY 101 DYIGSCTFEHFCVDLMDLIPGECPEPLRTYGLPCHGCFKEGTYSLPKSEF 152
    |||||
Db 101 DYIGSCTFEHFCVDLMDLIPGECPEPLRTYGLPCHGCFKEGTYSLPKSEF 152
    |||||

RESULT 2
S13195
ganglioside M2 activator protein - human
C:Species: Homo sapiens (man)
C>Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
R:Fuerst, W.; Schubert, J.; Machleidt, W.; Meyer, H.E.; Sandhoff, K.
R:Biochem. J. 294, 709-714, 1993
A:Title: The complete amino-acid sequences of human ganglioside GM2 activator protein and
A:Reference number: S13195; MUID:91006165; PMID:2209618
A:Accession: S13195
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-162 <FUB>
A:Cross-references: UNIPROT:P17900

Query Match      43.0%; Score 83; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 4.1e-77;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 GSTSVPLSSPLKVDLVLEKEVAGLWKIPCTDYIGSCTFEHFCVDLMDLIPGECPEPL 129
    |||||
Db 39 GSTSVPLSSPLKVDLVLEKEVAGLWKIPCTDYIGSCTFEHFCVDLMDLIPGECPEPL 98
    |||||

QY 130 RTYGLPCHGCFKEGTYSLPKSEF 152
    |||||
Db 99 RTYGLPCHGCFKEGTYSLPKSEF 121
    |||||

RESULT 3
S22411
ganglioside M2 activator protein (clone pGAP2) - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
C:Accession: S22411; S17108
R:Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
R:Biochem. J. 282, 807-813, 1992
A:Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A:Reference number: S22410; MUID:92207171; PMID:1554364
A:Accession: S22411
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-200 <NAG>
A:Cross-references: UNIPROT:P17900; EMBL:X61095; NID:g31856; PIDN:CAA43408.1; PID:g31857

Query Match      43.0%; Score 83; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 5e-77;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 GSTSVPLSSPLKVDLVLEKEVAGLWKIPCTDYIGSCTFEHFCVDLMDLIPGECPEPL 129
    |||||
Db 77 GSTSVPLSSPLKVDLVLEKEVAGLWKIPCTDYIGSCTFEHFCVDLMDLIPGECPEPL 136
    |||||

QY 130 RTYGLPCHGCFKEGTYSLPKSEF 152
    |||||
Db 137 RTYGLPCHGCFKEGTYSLPKSEF 159
    |||||

RESULT 4
S22411
ganglioside M2 activator protein precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 07-May-1999
C:Accession: S35613
R:Bellichioni, G.; Stirling, J.L.; Oriacchio, A.; Beccari, T.
R:Biochem. J. 294, 227-230, 1993
A:Title: Cloning and sequence analysis of a cDNA clone coding for the mouse G(M2) activa
A:Reference number: S35613; MUID:93371367; PMID:7689829
A:Accession: S35613
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-193 <BEL>

Query Match      9.3%; Score 18; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.0e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHGCFKEGTYSLP 148
    |||||
Db 131 TYGLPCHGCFKEGTYSLP 148
    |||||

RESULT 5
G95243
hypothetical protein SP2084 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: G95243
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
R:Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95243
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-291 <KUR>
A:Cross-references: UNIPROT:Q9X4T0; GB:AE005672; PIDN:AAK76144.1; PID:g14973594; GSPDB:G
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP2084

Query Match      4.7%; Score 9; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LSVVGSTSV 74
    |||||
Db 184 LSVVGSTSV 192
    |||||

RESULT 6
D98108
hypothetical protein pntS [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: D98108
R:Hoskins, J.A.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98108
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-312 <KUR>
A:Cross-references: UNIPROT:Q8DN64; GB:AE007317; PIDN:AAL00697.1; PID:g15459589; GSPDB:G
```


A;Gene: psts

Query Match 4.7%; Score 9; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 LSVVGSTSV 74
|||||
Db 205 LSVVGSTSV 213

RESULT 7
S56005
lysosomal protein 22K - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 07-May-1999
C;Accession: S56005
R;Kuwana, T.; Mullock, B.M.; Luzio, J.P.
Biochem. J. 308, 937-946, 1995
A;Title: Identification of a lysosomal protein causing lipid transfer, using a fluorescence
A;Reference number: S56005; MUID:97104296; PMID:8948454
A;Accession: S56005
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-20 <KUW>

Query Match 4.1%; Score 8; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 EGKDPAVI 48
|||||
Db 8 EGKDPAVI 15

RESULT 8
B75402
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: B75402
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75402
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-163 <WHI>
A;Cross-references: UNIPROT:Q9RUK1; GB:AE001984; GB:AE000513; NID:96459135; PIDN:AAF1096
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1385
A;Map position: 1

Query Match 4.1%; Score 8; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 ALGILLAT 19
|||||
Db 6 ALGILLAT 13

RESULT 9
S49337
fedF protein - Escherichia coli
C;Species: Escherichia coli
C;Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004
C;Accession: S49337
R;Imberechts, H.; Wild, P.; van Pelt, N.; van Montagu, M.; Charlier, G.; de Greve, H.; I

submitted to the EMBL Data Library, September 1993

A;Description: Characterisation of F107 fimbrial genes fedF and fedF conferring adhesion
A;Reference number: S49336
A;Accession: S49337
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-300 <IMB>
A;Cross-references: UNIPROT:Q47212; EMBL:Z26520; NID:g556632; PIDN:CAA81288.1; PID:g55663;
C;Superfamily: Escherichia coli fedF protein

Query Match 4.1%; Score 8; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 61 PGNVTLVS 68
|||||
Db 286 PGNVTLVS 293

RESULT 10
B81396
probable aminotransferase (degt family) Cj0505c [imported] - Campylobacter jejuni (strain
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: B81396
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hype
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: B81396
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-357 <PAR>
A;Cross-references: UNIPROT:Q9PI06; GB:AL139075; GB:AL111168; NID:96967817; PIDN:CAB7514;
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0505c
C;Superfamily: erythromycin resistance protein

Query Match 4.1%; Score 8; DB 2; Length 357;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 IESVLSSS 177
|||||
Db 21 IESVLSSS 28

RESULT 11
A40809
enamelin, 44K - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Apr-1992 #sequence_revision 10-Apr-1992 #text_change 09-Jul-2004
C;Accession: A40809
R;Deusch, D.; Palmon, A.; Fisher, L.W.; Kolodny, N.; Termine, J.D.; Young, M.F.
J. Biol. Chem. 266, 16021-16028, 1991
A;Title: Sequencing of bovine enamel ("tuftelin") a novel acidic enamel protein.
A;Reference number: A40809; MUID:91340750; PMID:1874744
A;Accession: A40809
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-389 <DEU>
A;Cross-references: UNIPROT:P27628; GB:M64924; NID:g163794; PIDN:AAA30797.1; PID:g163795
C;Keywords: phosphoprotein

Query Match 4.1%; Score 8; DB 2; Length 389;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 LEKEVAGL 93
|||||
Db 225 LEKEVAGL 232

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: H83117

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-574 <STO>

A:Cross-references: UNIPROT:Q9HWG7; GB:AE004839; GB:AE004091; NID:g9950434; PIDN:AAG0761

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA4222

Query Match 4.1%; Score 8; DB 2; Length 574;

Best Local Similarity 100.0%; Pred. No. 9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TALLGULLA 18

Db 154 TALLGULLA 161

RESULT 15

T04223

hypothetical protein T5C23.190 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004

C:Accession: T04223

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, B.; Brandt, A.; Duesterhoeft, A.; Bancroft,

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15261

A:Accession: T04223

A:Molecule type: DNA

A:Residues: 1-96 <BEV>

A:Cross-references: UNIPROT:Q9T0E3; EMBL:AL049500

A:Experimental source: cultivar Columbia; BAC clone T5C23

C:Genetics:

A:Map position: 4

A:Introns: 21/1

A:Note: T5C23.190

Query Match 3.6%; Score 7; DB 2; Length 96;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKPSQLS 32

Db 2 KKPSQLS 8

Search completed: November 17, 2004, 09:38:40

Job time : 6.45839 secs

RESULT 12

B82984

hypothetical protein PA5295 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C:Accession: B82984

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: B82984

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-558 <STO>

A:Cross-references: UNIPROT:Q9HTQ9; GB:AE004942; GB:AE004091; NID:g9951607; PIDN:AAG0868

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA5295

Query Match 4.1%; Score 8; DB 2; Length 558;

Best Local Similarity 100.0%; Pred. No. 8.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GLLLATPA 21

Db 55 GLLLATPA 62

RESULT 13

AH2321

hypothetical protein all4127 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C:Accession: AH2321

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AH2321

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-564 <KUR>

A:Cross-references: UNIPROT:Q8YPR5; GB:BA000019; PIDN:BA875826.1; PID:gl7133262; GSPDB:G

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all4127

Query Match 4.1%; Score 8; DB 2; Length 564;

Best Local Similarity 100.0%; Pred. No. 8.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PLLIALGL 15

Db 359 PLLIALGL 366

RESULT 14

H83117

probable ATP-binding component of ABC transporter PA4222 [imported] - Pseudomonas aerugi

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C:Accession: H83117

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 07:28:04 ; Search time 32.1224 Seconds
(without alignments)
3457.007 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 193

Sequence: 1 MSLMQAPLLIALGLLALTP.....LSSSGKRLGCIKTAASLKGI 193

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Uniprot 02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	58.0	189	2	Q6LBL5
2	112	58.0	189	2	CAA43994
3	112	58.0	193	1	SAP3_HUMAN
4	100	51.8	103	2	Q14427
5	28	14.5	190	2	Q8HXX6
6	19	9.8	197	2	Q6GLN6
7	18	9.3	47	2	Q8I028
8	18	9.3	193	1	SAP3_MOUSE
9	18	9.3	199	2	Q6IN37
10	18	9.3	199	2	Q8CUH4
11	11	5.7	146	2	Q6Q7X4
12	11	5.7	146	2	AA564351
13	11	5.7	151	2	Q6Q7X5
14	11	5.7	151	2	AA564350
15	10	5.2	721	2	Q8ALY3
16	9	4.7	291	2	Q9X4T0
17	9	4.7	312	2	Q8DM64
18	9	4.7	401	2	Q88VX3
19	8	4.1	20	2	Q9QWJ2
20	8	4.1	26	2	Q8I029
21	8	4.1	163	2	Q9RUK1
22	8	4.1	184	2	Q6DT75
23	8	4.1	188	1	TFT1_PIG
24	8	4.1	257	1	PPNK_RHIME
25	8	4.1	300	2	Q47212
26	8	4.1	302	2	Q7P200
27	8	4.1	323	2	Q9SHU6
28	8	4.1	349	2	Q9XN28
29	8	4.1	357	2	Q9PI06
30	8	4.1	365	2	BAC11346
31	8	4.1	377	2	Q6NBZ3

32	8	4.1	377	2	CAE26127
33	8	4.1	378	2	Q8HYU6
34	8	4.1	390	1	TFT1_BOVIN
35	8	4.1	390	1	TFT1_HUMAN
36	8	4.1	414	1	CBX5_MOUSE
37	8	4.1	444	2	Q6MWL1
38	8	4.1	444	2	CAE76287
39	8	4.1	447	2	Q6K5N1
40	8	4.1	447	2	BAD22094
41	8	4.1	448	2	Q95DW6
42	8	4.1	467	2	Q762L6
43	8	4.1	467	2	BAD03942
44	8	4.1	511	2	Q8RAZ4
45	8	4.1	525	2	Q7VCD6

Cae26127 rhodopsin
Q8hyu6 oryctolagus
P27628 bos taurus
Q9nnx1 homo sapien
Q9dbv5 mus musculus
Q8mvl1 neurospora
Cae76287 neurospora
Q6k5n1 oryza sativ
Bad22094 oryza sat
Q95dw6 goodenia ov
Q762l6 klebsiella
Bad03942 klebsiell
Q8raz4 thermoaer
Q7vcd6 prochloroco

ALIGNMENTS

RESULT 1

ID	Q6LBL5	PRELIMINARY;	PRT;	189 AA.
AC	Q6LBL5			
DT	05-JUL-2004	(TrEMBLrel. 27, Created)		
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)		
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)		
DE	GM2 activator protein.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92008637; PubMed=1915857;			
RA	Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,			
RA	Sandhoff K.;			
RT	"Characterization of full-length cDNAs and the gene coding for the			
RT	human GM2 activator protein.";			
RL	FEBS Lett. 289:260-264(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93277527; PubMed=8503891;			
RA	Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,			
RA	Sandhoff K.;			
RT	"Over-expression of a functionally active human GM2-activator protein			
RT	in escherichia coli.";			
RL	Biochem. J. 292:571-576(1993).			
DR	EMBL; X62078; CAA43994.1;			
DR	InterPro; IPR003172; EI_DerP2_DerF2.			
DR	SMART; SM00737; MU; 1.			
SQ	SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;			

Query Match 58.0%; Score 112; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.6e-104;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	41	EGKDPVIRSLTLEPDPVIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPT	100
Db	37	EGKDPVIRSLTLEPDPVIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPT	96
QY	101	DYIGSCTFHFCDVLDMLIPTGPEPCPEPLRTVGLPCHCFKGTYSLPKSEF	152
Db	97	DYIGSCTFHFCDVLDMLIPTGPEPCPEPLRTVGLPCHCFKGTYSLPKSEF	148

RESULT 2

ID	CAA43994	PRELIMINARY;	PRT;	189 AA.
AC	CAA43994			
DT	02-MAR-2004	(TrEMBLrel. 27, Created)		
DT	02-MAR-2004	(TrEMBLrel. 27, Last sequence update)		
DT	02-MAR-2004	(TrEMBLrel. 27, Last annotation update)		
DE	GM2 activator protein.			

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
RT human GM2 activator protein.";
RL FEBS Lett. 289:260-264(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277527; PubMed=8503891;
RA Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA Sandhoff K.;
RT "Over-expression of a functionally active human GM2-activator protein
RT in Escherichia coli.";
RL Biochem. J. 292:571-576(1993).
DR EMBL; X62078; CAA43994.1; -
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match 58.0%; Score 112; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.6e-104; Mismatches 0; Gaps 0;
Matches 112; Conservative 0; Indels 0;

QY 41 EKGDAVIRSLTLEPDPVWPGNVTLVVGSTSVPLSKVDLVLEKEVAGLWIKIPT 100
Db |||
37 EKGDAVIRSLTLEPDPVWPGNVTLVVGSTSVPLSKVDLVLEKEVAGLWIKIPT 96
|||

QY 101 DYIGCTEHHFCDVLDMLPTGCPPEPLRTYGLPCHCPFKGTYSLPKSEF 152
|||
97 DYIGCTEHHFCDVLDMLPTGCPPEPLRTYGLPCHCPFKGTYSLPKSEF 148
|||

RESULT 3
SAP3 HUMAN
ID SAP3 HUMAN STANDARD; PRT; 193 AA.
AC P17300; Q14426; Q14428;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate
DE activator protein) (Shingolipid activator protein 3) (SAP-3).
GN Name=GM2A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RX MEDLINE=91282768; PubMed=2059210;
RA Xie B., McInnes B., Neote K., Lambonwah A.-M., Mahuran D.;
RT "Isolation and expression of a full-length cDNA encoding the human G-
RT M2 activator protein.";
RL Biochem. Biophys. Res. Commun. 177:1217-1223(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
RT human GM2 activator protein.";
RL FEBS Lett. 289:260-264(1991).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RC TISSUE=Placenta;
RX MEDLINE=92207171; PubMed=1554364;
RA Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
RT "Evidence for two cDNAs encoding human GM2-activator protein.";
RL Biochem. J. 282:807-813(1992).
RN [4]

RP SEQUENCE FROM N.A.
RX MEDLINE=93052421; PubMed=1427911;
RA Xie B., Kennedy J.L., McInnes B., Auger D., Mahuran D.J.;
RT "Identification of a processed pseudogene related to the functional
RT gene encoding the GM2 activator protein: localization of the
RT pseudogene to human chromosome 3 and the functional gene to human
RT chromosome 5.";
RL Chromosomes 14:796-798(1992).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=9294584; PubMed=10364519;
RA Chen B., Rigat B., Curry C., Mahuran D.J.;
RT "Structure of the GM2A gene: identification of an exon 2 nonsense
RT mutation and a naturally occurring transcript with an in-frame
RT deletion of exon 2.";
RL Am. J. Hum. Genet. 65:77-87(1999).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
RA Bosak S.A., Morley P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Skailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [7]
RP SEQUENCE OF 15-193 FROM N.A.
RX MEDLINE=89325664; PubMed=2753159;
RA Schroeder M., Klima H., Nakano T., Kwon H., Quintern L.E.,
RA Gaertner S., Suzuki K., Sandhoff K.;
RT "Isolation of a cDNA encoding the human GM2 activator protein.";
RL FEBS Lett. 251:197-200(1989).
RN [8]
RP SEQUENCE OF 32-193.
RC TISSUE=Kidney;
RX MEDLINE=91006165; PubMed=2209618;
RA Furst W., Schubert J., Machleidt W., Meyer H.E., Sandhoff K.;
RT "The complete amino-acid sequences of human ganglioside GM2 activator
RT protein and cerebroside sulfate activator protein.";
RL Eur. J. Biochem. 192:709-714(1990).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=20545600; PubMed=11090283; DOI=10.1006/jmbi.2000.4225;
RA Wright C.S., Li S.-C., Rastinejad F.;
RT "Crystal structure of human GM2-activator protein with a novel beta-
RT cup topology.";
RL J. Mol. Biol. 304:411-422(2000).
RN [10]
RP VARIANT TSD-AB ARG-138.
RX MEDLINE=92008638; PubMed=1915858;
RA Schroeder M., Schnabel D., Suzuki K., Sandhoff K.;
RT "A mutation in the gene of a glycolipid-binding protein (GM2
RT activator) that causes GM2-gangliosidosis variant AB.";
RL FEBS Lett. 290:1-3(1991).
RN [11]
RP VARIANT TSD-AB PRO-169.
RX MEDLINE=94063850; PubMed=8244332;
RA Schroeder M., Schnabel D., Hurwitz R., Young E., Suzuki K.,

```

RA Sandhoff K.;
 RT "Molecular genetics of GM2-gangliosidosis AB variant: a novel mutation
 RL and expression in BHK cells."; Hum. Genet. 92:437-440(1993).
 RN [12]
 RP VARIANT TSD-AB LYS-88 DEL.
 RX MEDLINE=97055887; PubMed=8900233;
 RA Schepers U., Glombitza G., Lemm T., Hoffmann A., Chabas A., Ozand P.,
 RA Sandhoff K.;
 RT "Molecular analysis of a GM2-activator deficiency in two patients with
 RL GM2-gangliosidosis AB variant."; Am. J. Hum. Genet. 59:1048-1056(1996).
 CC -!- FUNCTION: Binds gangliosides and stimulates ganglioside GM2
 CC degradation. It stimulates only the breakdown of ganglioside GM2
 CC and glycolipid GA2 by beta-hexosaminidase A. It extracts single
 CC GM2 molecules from membranes and presents them in soluble form to
 CC beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and
 CC conversion to GM3.
 CC -!- SUBCELLULAR LOCATION: Lysosomal.
 CC -!- DISEASE: Defects in GM2A are the cause of Tay-Sachs disease AB
 CC variant (TSD-AB) [MIM:272750]; also known as GM2-gangliosidosis
 CC type AB.
 CC -!- DATABASE: NAME=GM2Adb; NOTE=GM2A mutation database;
 CC WWW="http://www.hexdb.mcgill.ca/?Topic=GM2Adb&Page=MutationSubmission".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M76477; AAA35907.1; -;
 DR EMBL; X62078; CAA43993.1; -;
 DR EMBL; X61095; CAA43408.1; ALT_INIT.
 DR EMBL; L01439; AAA52767.1; -;
 DR EMBL; AF124719; AAD25741.1; -;
 DR EMBL; AF124717; AAD25741.1; JOINED.
 DR EMBL; AF124718; AAD25741.1; JOINED.
 DR EMBL; BC009273; AAH09273.1; -;
 DR EMBL; X16087; CAA34215.1; -;
 DR PIR; I54178; I54178.
 DR PIR; S13195; S13195.
 DR PIR; S22411; S22411.
 DR PDB; 1G13; X-ray; A/B/C=32-193.
 DR Genew; HGNC:4367; GM2A.
 DR MIM; 272750; -;
 DR GO; GO:0005764; C:lysosome; NAS.
 DR GO; GO:0030290; F:sphingolipid activator protein activity; NAS.
 DR GO; GO:0019377; P:glycolipid catabolism; NAS.
 DR GO; GO:0030149; P:sphingolipid catabolism; NAS.
 DR InterPro; IPR003172; E1_DerP2_DerF2.
 DR SMART; SM00737; ML; 1.
 KW 3D-structure; Direct protein sequencing; Disease mutation;
 KW Glycoprotein; GM2-gangliosidosis; Lysosome; Polymorphism; Signal;
 KW Sphingolipid metabolism.
 FT SIGNAL 1 31
 FT CHAIN 32 193 Ganglioside GM2 activator.
 FT DISULFID 39 183
 FT DISULFID 99 106
 FT DISULFID 112 138
 FT DISULFID 125 136
 FT CARBOHYD 63 63 N-linked (GlcNAc...).
 FT VARIANT 19 19 T -> A.
 FT VARIANT 32 33 Missing (in 80% of the protein).
 FT VARIANT 88 88 Missing (in TSD-AB).
 FT VARIANT 138 138 Missing (in TSD-AB).
 FT VARIANT 169 169 C -> R (in TSD-AB).
 FT VARIANT 169 169 R -> P (in TSD-AB).

FT FT
 FT CONFLICT 59 59 /FTId=VAR_011698.
 FT CONFLICT 69 69 V -> I (in Ref. 3).
 FT STRAND 35 38 V -> M (in Ref. 3).
 FT TURN 41 43
 FT STRAND 46 54
 FT STRAND 58 59
 FT STRAND 63 72
 FT STRAND 76 76
 FT STRAND 81 90
 FT TURN 91 92
 FT STRAND 93 96
 FT STRAND 100 100
 FT TURN 101 102
 FT STRAND 103 103
 FT STRAND 107 108
 FT TURN 109 110
 FT HELIX 111 118
 FT TURN 121 122
 Query Match 58.0%; Score 112; DB 1; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.6e-104;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 EGKDPAVIRSLTLEPDPPIVPGNVTLSSPLKVDLVLEKEVAGLWIKIPCT 100
 DB 41 EGKDPAVIRSLTLEPDPPIVPGNVTLSSPLKVDLVLEKEVAGLWIKIPCT 100
 QY 101 DYIGSTCFEHFCDVLDMLIPTGECPCPEPLRTYGLPCHCPKCTYSLPKSEF 152
 DB 101 DYIGSTCFEHFCDVLDMLIPTGECPCPEPLRTYGLPCHCPKCTYSLPKSEF 152
 RESULT 4
 Q14427 PRELIMINARY; PRT; 103 AA.
 ID Q14427
 AC Q14427
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE GM2-activator protein (Fragment).
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=922071171; PubMed=1554364;
 RA Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
 RT "Evidence for two cDNAs encoding human GM2-activator protein.";
 RL Biochem. J. 282:807-813(1992).
 DR EMBL; X61094; CAA43407.1; -;
 DR NON_TER 1
 SQ SEQUENCE 103 AA; 11137 MW; 434BFC334E976F91 CRC64;
 Query Match 51.8%; Score 100; DB 2; Length 103;
 Best Local Similarity 100.0%; Pred. No. 1.2e-92;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 43 KDPAVIRSLTLEPDPPIVPGNVTLSSPLKVDLVLEKEVAGLWIKIPCTDY 102
 DB 1 KDPAVIRSLTLEPDPPIVPGNVTLSSPLKVDLVLEKEVAGLWIKIPCTDY 102
 QY 103 IGSCTFEHFCDVLDMLIPTGECPCPEPLRTYGLPCHCPKKE 142
 DB 61 IGSCTFEHFCDVLDMLIPTGECPCPEPLRTYGLPCHCPKKE 100
 RESULT 5
 Q8HXX6 PRELIMINARY; PRT; 190 AA.
 ID Q8HXX6
 AC Q8HXX6;

DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Ganglioside GM2 activator.
 GN Name=gm2a;
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain cerebellum cortex;
 RA Kusuda J., Osada N., Hida M., Sugano S., Hashimoto K.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB083313; BAC20592.1; -;
 DR HSSP; P17900; IG13.
 DR InterPro; IPR003172; El_DerP2_DerP2.
 DR SMART; SM00737; ML; 1.
 SQ SEQUENCE 190 AA; 20494 MW; 9F9582BEB75715C3 CRC64;

 Query Match 14.5%; Score 28; DB 2; Length 190;
 Best Local Similarity 100.0%; Pred. No. 1.5e-19;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 125 CPEPLRTYGLPCHCPKSGTSLPKSEF 152
 |||||
 DB 122 CPEPLRTYGLPCHCPKSGTSLPKSEF 149

 RESULT 6
 Q6GLN6 PRELIMINARY; PRT; 197 AA.
 AC Q6GLN6
 DT 05-JUL-2004 (Tremblrel. 27, Created)
 DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schermer A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 RX MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;

RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 initiative.";
 RT Dev. Dyn. 225:384-391 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 RA Klein S., Strausberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC074424; AAH74424.1; -;
 DR InterPro; IPR003172; El_DerP2_DerP2.
 DR SMART; SM00737; ML; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 197 AA; 21401 MW; A0909CAD82FD66EB CRC64;

 Query Match 9.8%; Score 19; DB 2; Length 197;
 Best Local Similarity 100.0%; Pred. No. 2e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 125 CPEPLRTYGLPCHCPKSGTSLPKSEF 143
 |||||
 DB 125 CPEPLRTYGLPCHCPKSGTSLPKSEF 143

 RESULT 7
 Q81028 PRELIMINARY; PRT; 47 AA.
 AC Q81028
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Ganglioside GM2 activator (Fragment).
 GN Name=gm2a;
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kusuda J., Osada N., Hashimoto K.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB083313; BAC20643.1; -;
 FT NON_TER 1
 FT NON_TER 47
 SQ SEQUENCE 47 AA; 4883 MW; C5F5537F3A029FEB CRC64;

 Query Match 9.3%; Score 18; DB 2; Length 47;
 Best Local Similarity 100.0%; Pred. No. 5.3e-10;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 41 EGKDPAVIRSLTLEDDPI 58
 |||||
 DB 11 EGKDPAVIRSLTLEDDPI 28

 RESULT 8
 SAP3_MOUSE STANDARD; PRT; 193 AA.
 ID SAP3_MOUSE
 AC Q60648; Q61610; Q61819;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate
 activator protein) (Shingolipid activator protein 3) (SAP-3).
 GN Name=Gm2a;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
 RX MEDLINE=95229165; PubMed=7713516;

Yamanaka S., Johnson O.N., Lyu M.S., Kozak C.A., Proia R.L.:
 "The mouse gene encoding the GM2 activator protein (Gm2a): cDNA
 sequence, expression, and chromosome mapping.";
 Genomics 24:601-604 (1994).
 [2]
 MEDLINE=93371367; PubMed=7689829;
 Bellachio G., Stirling J.L., Orlacchio A., Beccari T.;
 "Cloning and sequence analysis of a cDNA clone coding for the mouse
 GM2 activator protein.";
 Biochem. J. 294:227-230 (1993).
 [3]
 SEQUENCE FROM N.A.
 STRAIN=C57BL/6 X CBA;
 MEDLINE=97224573; PubMed=9060405;
 Bertoni C., Appolloni M.G., Stirling J.L., Li S.C., Li Y.T.,
 Orlacchio A., Beccari T.;
 "Structural organization and expression of the gene for the mouse GM2
 activator protein.";
 Mamm. Genome 8:90-93 (1997).
 [4]
 SEQUENCE FROM N.A.
 STRAIN=C57BL/6J; TISSUE=Mammary gland;
 MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 Butterfield Y.S., Grimwood J., Schmutz J., Myers R.M., Smailus D.E.,
 Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [1]
 FUNCTION: Binds gangliosides and stimulates ganglioside GM2
 degradation. It stimulates only the breakdown of ganglioside GM2
 and glycolipid GA2 by beta-hexosaminidase A. It extracts single
 GM2 molecules from membranes and presents them in soluble form to
 beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and
 conversion to GM3.
 [2]
 SUBCELLULAR LOCATION: Lysosomal.
 [3]
 TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney and
 testis.
 [4]
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; U09816; AAA21543.1; -;
 EMBL; L19526; AAA61929.1; -;
 EMBL; U34359; AAB06275.1; ALT SEQ.
 EMBL; U34356; AAB06275.1; JOINED.
 EMBL; U34357; AAB06275.1; JOINED.
 EMBL; U34358; AAB06275.1; JOINED.
 EMBL; BC004651; AA04651.1; -;
 EMBL; P17900; IG13.
 DR MGD; MGI:95762; Gm2a.
 KW Glycoprotein; Lysosome; Signal; sphingolipid metabolism.
 FT SIGNAL 1 31 By similarity.

FT CHAIN 32 193 Ganglioside GM2 activator.
 FT DISULFID 39 183 By similarity.
 FT DISULFID 99 106 By similarity.
 FT DISULFID 112 138 By similarity.
 FT DISULFID 125 136 By similarity.
 FT CARBOHYD 151 151 N-linked (GlcNAc...) (Potential).
 FT CONFLICT 53 I -> T (in Ref. 1).
 SQ SEQUENCE 193 AA; 20824 MW; 59CC4ABE56FA1FC7 CRC64;
 Query Match 9.3%; Score 18; DB 1; Length 193;
 Best Local Similarity 100.0%; Pred. No. 2e-09;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 131 TYGLPCHCPFKGTYSLP 148
 Db 131 TYGLPCHCPFKGTYSLP 148
 RESULT 9
 Q6IN37 PRELIMINARY; PRT; 199 AA.
 AC Q6IN37;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE GM2 ganglioside activator protein.
 GN Name=Gm2a;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 Butterfield Y.S., Grimwood J., Schmutz J., Myers R.M., Smailus D.E.,
 Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [2]
 SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RA Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC072474; AAH72474.1; -;
 DR InterPro; IPR003172; El_DerP2_DerF2.
 DR SMART; SM00737; ML; 1.
 SQ SEQUENCE 199 AA; 21493 MW; C2B5203FDFABF507 CRC64;
 Query Match 9.3%; Score 18; DB 2; Length 199;
 Best Local Similarity 100.0%; Pred. No. 2e-09;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 131 TYGLPCHCPFKGTYSLP 148
 Db 137 TYGLPCHCPFKGTYSLP 154

```
DE GM2 activator protein variant AB (Fragment).
GN GM2A.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
RA Dodson A.N., Gentry A.S., Griffin B., Rolfsma M.D., Baker H.J.;
RT "Mutation of the GM2 activator protein in a feline model of GM2
RT gangliosidosis."
RT Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY553654; AAS64351.1; -.
DR NON_TER 1
FT SEQUENCE 146 AA; 15820 MW; 754E06C0CFCFA74E3 CRC64;

Query Match 5.7%; Score 11; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPFK 141
Db 89 TYGLPCHCPFK 99

RESULT 13
Q6Q7X5 PRELIMINARY; PRT; 151 AA.
AC Q6Q7X5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE GM2 activator protein (Fragment).
GN Name=GM2A;
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
RA Dodson A.N., Gentry A.S., Griffin B., Rolfsma M.D., Baker H.J.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY553653; AAS64350.1; -.
DR InterPro; IPR003172; EI_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
FT NON_TER 1
FT SEQUENCE 151 AA; 16310 MW; E23EF16CE5844A5F CRC64;

Query Match 5.7%; Score 11; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPFK 141
Db 89 TYGLPCHCPFK 99

RESULT 14
AAS64350 PRELIMINARY; PRT; 151 AA.
AC AAS64350;
DT 25-MAR-2004 (TrEMBLrel. 27, Created)
DT 25-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE GM2 activator protein (Fragment).
GN GM2A.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
```

```
DE GM2 activator protein variant AB (Fragment).
GN GM2A.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
RA Dodson A.N., Gentry A.S., Griffin B., Rolfsma M.D., Baker H.J.;
RT "Mutation of the GM2 activator protein in a feline model of GM2
RT gangliosidosis."
RT Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY553654; AAS64351.1; -.
DR NON_TER 1
FT SEQUENCE 146 AA; 15820 MW; 754E06C0CFCFA74E3 CRC64;

Query Match 9.3%; Score 18; DB 2; Length 199;
Best Local Similarity 100.0%; Pred. No. 2e-09; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPFKGTYSLP 148
Db 137 TYGLPCHCPFKGTYSLP 154

RESULT 11
Q6Q7X4 PRELIMINARY; PRT; 146 AA.
AC Q6Q7X4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE GM2 activator protein variant AB (Fragment).
GN Name=GM2A;
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
RA Dodson A.N., Gentry A.S., Griffin B., Rolfsma M.D., Baker H.J.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY553654; AAS64351.1; -.
DR InterPro; IPR003172; EI_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
FT NON_TER 1
FT SEQUENCE 146 AA; 15820 MW; 754E06C0CFCFA74E3 CRC64;

Query Match 5.7%; Score 11; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPFK 141
Db 89 TYGLPCHCPFK 99

RESULT 12
AAS64351 PRELIMINARY; PRT; 146 AA.
AC AAS64351;
DT 25-MAR-2004 (TrEMBLrel. 27, Created)
DT 25-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-MAR-2004 (TrEMBLrel. 27, Last annotation update)
```


RP SEQUENCE FROM N.A.
RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
RT Dodson A.N., Gentry A.S., Griffin B., Rolfsma M.D., Baker H.J.;
RT "Mutation of the GM2 activator protein in a feline model of GM2
RT gangliosidosis."
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY553653; AAS64350.1; -.
FT NON_TER 1
SQ SEQUENCE 151 AA; 16310 MW; E23EF16CE5844A5F CRC64;

Query Match 5.7%; Score 11; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TYGLPCHCPFK 141
|||
DB 89 TYGLPCHCPFK 99

RESULT 15

Q8AIY3 PRELIMINARY; PRT; 721 AA.
AC Q8AIY3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocNames=BT3525;
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RT Chiang H.C., Hooper L.V., Gordon J.I.;
RL "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."
DR Science 299:2074-2076(2003).
DR EMBL; AR016941; AAO78631.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 721 AA; 83162 MW; 9F02F07EAA991182 CRC64;

Query Match 5.2%; Score 10; DB 2; Length 721;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LLIATGLLLA 18
|||
DB 6 LLIATGLLLA 15

Search completed: November 17, 2004, 09:34:48
Job time : 32.1224 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 19, 2004, 00:47:47 ; Search time 4616 Seconds
(without alignments)
1977.234 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 193

Sequence: 1 MQLMQAPLLIALGLLALTP.....LSSSGKRLGCIKIAASLKGI 193

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPTO.spool.p/US10030937/runat_16112004_153102_3286/app.query.fasta.1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=olip2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937 @CNC_1.1.7406 @runat_16112004_153102_3286 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPOP=60
-FGAPEXT=7 -YGAPOP=60 -YGAPOP=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

RESULT 1
HSGM2AP
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED

COMMENT
FEATURES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	58.0	821	9	X16087 Human mRNA
2	112	58.0	953	9	M76477 Human G-M2
3	112	58.0	1043	9	L01439 Human GM2-a
4	112	58.0	2413	9	BC009273 Homo sapi

5	112	58.0	2436	6	AX330938
6	112	58.0	2436	9	HSGM2APT
7	100	51.8	648	9	HSGM2ADA
8	83	43.0	1045	6	CQ728078 Sequence
9	83	43.0	1093	9	HSGM2APB
10	61	31.6	475	6	CQ050684 Sequence
11	61	31.6	475	6	CQ065722 Sequence
12	61	31.6	475	6	CQ092707 Sequence
13	61	31.6	475	6	CQ131515 Sequence
14	61	31.6	475	6	CQ170122 Sequence
15	61	31.6	475	6	CQ199273 Sequence
16	61	31.6	475	6	CQ214720 Sequence
17	61	31.6	475	6	CQ253322 Sequence
18	61	31.6	475	6	CQ290453 Sequence
19	61	31.6	475	6	CQ327432 Sequence
20	61	31.6	1706	9	HSGM2A3
21	61	31.6	107320	2	AC011391
22	61	31.6	120594	9	AC011342
23	61	31.6	151712	9	AC008385
24	48	24.9	145	6	CQ055682 Sequence
25	48	24.9	145	6	CQ074936 Sequence
26	48	24.9	145	6	CQ105860 Sequence
27	48	24.9	145	6	CQ144548 Sequence
28	48	24.9	145	6	CQ180027 Sequence
29	48	24.9	145	6	CQ204401 Sequence
30	48	24.9	145	6	CQ227751 Sequence
31	48	24.9	145	6	CQ265915 Sequence
32	48	24.9	145	6	CQ302985 Sequence
33	48	24.9	145	6	CQ340179 Sequence
34	42	21.8	1047	9	HSGM2A2
35	40	20.7	352	11	G05651
36	28	14.5	950	9	AB083313
37	27	14.0	633	9	HSGM2A1
38	19	9.8	3684	5	BC074424
39	18	9.3	257	9	AB083332
40	18	9.3	600	10	AB051391
41	18	9.3	1113	10	MUSGM2ACT
42	18	9.3	1983	6	E12286
43	18	9.3	1983	6	AX827433
44	18	9.3	2003	10	MMU09816
45	18	9.3	2024	10	BC004651

ALIGNMENTS

HSGM2AP 821 bp mRNA linear PRI 19-JUL-1995
Human mRNA for G(M2) activator protein.
X16087
X16087.1 GI:31852
G(M2) activator protein; G(M2) gangliosidosis.
Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Schroder, M., Klima, H., Nakano, T., Kwon, H., Quintern, L. E., Gartner, S., Suzuki, K. and Sandhoff, K.

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

Location/Qualifiers

Over-expression of a functionally active human GM2-activator protein in Escherichia coli

Biochem. J. 292 (Pt 2), 571-576 (1993)

8503891

Data kindly reviewed (23-JAN-1991) by Sandhoff K.

```

source
1. .821
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="pGAP1"
/cell_type="fibroblast"
<1. .540
/contig_start=1
/product="G(M2) activator protein"
/protein_id="CAA34215.1"
/db_xref="GI:31853"
/db_xref="GOA:P17900"
/db_xref="Swiss-Prot:P17900"
/translation="LLLATPAQAHLLKPSQSSSWNCDRGKDPVARSITLPEPDI
VPGNVLTVSVGTSVPLSPKVDLVLEKEVAGLWIKIPCTYIGSCTFHFCDVLD
MLIPTGEPPEPLRTYGLPCHCPKPEKGTYSLPKSEFVVPDLELPSWLTGNYRIESVL
SSSGKRLGCIKIAASLKG1"
<1. .51
sig_peptide
mat_peptide
52. .537
/product="G(M2) activator protein"
/evidence=experimental

ORIGIN
Alignment Scores:
Pred. No.: 1,248-106 Length: 821
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservative: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 9 Gaps: 0

US-10-030-937-9 (1-193) x HSGM2AP (1-821)
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
DB 79 GAAGGAAGGACCTCGCGTGATCAGAGCCCTGACTCTGGAGCCTGACCCCATCGTCTG 138
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
DB 139 CTGGAAATGTGACCTCAGTGTCTGGGAGCAGCAGTGTCTCCCTGAGTTCTCTCTG 198
QY 81 LysValAspLeuValLeuGlyLysGluValAlaGlyLeuTyrIleLysIleProCysThr 100
DB 199 AAGTGGATTAGTTTGGAGAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGCACA 258
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
DB 259 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTCTTGACATGTTAATTCCT 318
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
DB 319 ACTGGGAGCCCTGCCAGAGCCCTCGGTACTATGGGCTTCTTGGCCACTGTCCCTTC 378
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
DB 379 AAAGAGGAGAACCTACTCACTGCCAAGAGCGAATTCGT-TGTGCTGACTGGAGCTGCC 437
QY 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
DB 438 CAGTTGGCTCACCACCGGAACTACCGCATAGAGCGTCTTGAGCAGCAGTGGGAAGCG 497
QY 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
DB 498 TCTGGGCTGCATCAAGATCGTGCCTCTCTAAAGGGCATA 537

RESULT 2
HUMGM2
LOCUS HUMGM2 953 bp mRNA linear PRI 27-APR-1993
DEFINITION Human G-M2 activator protein mRNA, complete cds.
ACCESSION M76477
VERSION M76477.1 GI:183356
KEYWORDS G-M2 activator protein.
SOURCE Homo sapiens (human)

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 953)
AUTHORS Xie,B., McInnes,B., Neote,K., Lambonwah,A.M. and Mahuran,D.
TITLE Isolation and expression of a full-length cDNA encoding the human
G-M2 activator protein
JOURNAL Biochem. Biophys. Res. Commun. 177 (3), 1217-1223 (1991)
MEDLINE 91282768
PUBMED 2059210
COMMENT
FEATURES
source
1. .953
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
91. .672
/codon_start=1
/product="G-M2 activator protein"
/protein_id="AAA35907.1"
/db_xref="GI:183357"
/translation="MQSLMQAPLLIALGALLAQAHLKPKSQSSSWNCDRGKDY
PAVIRSLTEPDIIVPGNVLTVSVGTSVPLSPKVDLVLEKEVAGLWIKIPCTDY
IGSCTFHFCDVLDMLIPTGEPPEPLRTYGLPCHCPKPEKGTYSLPKSEFVVPDLELP
SWLTGNYRIESVLSSSGKRLGCIKIAASLKG1"
91. .159
sig_peptide
mat_peptide
160. .669
/product="G-M2 activator protein"

ORIGIN
Alignment Scores:
Pred. No.: 1,428-106 Length: 953
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservative: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 9 Gaps: 0

US-10-030-937-9 (1-193) x HUMGM2 (1-953)
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
DB 211 GAAGGAAGGACCTCGCGTGATCAGAGCCCTGACTCTGGAGCCTGACCCCATCGTCTG 270
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
DB 271 CTGGAAATGTGACCTCAGTGTCTGGGAGCAGCAGTGTCTCCCTGAGTTCTCTCTG 330
QY 81 LysValAspLeuValLeuGlyLysGluValAlaGlyLeuTyrIleLysIleProCysThr 100
DB 331 AAGTGGATTAGTTTGGAGAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGCACA 390
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
DB 391 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTCTTGACATGTTAATTCCT 450
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
DB 451 ACTGGGAGCCCTGCCAGAGCCCTCGGTACTATGGGCTTCTTGGCCACTGTCCCTTC 510
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
DB 511 AAAGAGGAGAACCTACTCACTGCCAAGAGCGAATTCGT-TGTGCTGACTGGAGCTGCC 569
QY 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
DB 570 CAGTTGGCTCACCACCGGAACTACCGCATAGAGCGTCTTGAGCAGCAGTGGGAAGCG 629
QY 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
DB 630 TCTGGGCTGCATCAAGATCGTGCCTCTCTAAAGGGCATA 669

RESULT 3

```

HUMGM2A
LOCUS 1043 bp mRNA linear PRI 09-NOV-1994
DEFINITION Human GM2-activator protein (GM2A) mRNA, complete cds.
L01439
ACCESSION L01439.1 GI:183358
VERSION GM2 activator protein.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1043)
Xie, B., Kennedy, J.L., McInnes, B., Auger, D. and Mahuran, D.
Identification of a processed pseudogene related to the functional gene encoding the GM2 activator protein: localization of the pseudogene to human chromosome 3 and the functional gene to human chromosome 5
Genomics 14 (3), 796-798 (1992)
JOURNAL
MEDLINE 93052421
PUBMED 1427911
COMMENT Original source text: Homo sapiens cDNA to mRNA.
FEATURES
source Location/Qualifiers
1..1043
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="5"
1..1043
/genes="GM2A"
91..672
/genes="GM2A"
/codon_start=1
/product="GM2-activator protein"
/protein_id="AA52767.1"
/db_xref="GI:183359"
/translation="MOSLMQAPLLIALGLLAPQAHLKPKSLSFSWDCDEGK
PAVRLSLTEPPDIPVVGWNTLSVVGSTVPLSLKVDLVLEKEVAGLWIKPCTDY
IGSCFTEHCDVLMDLPTGEPCEPLRYGLPCHCFKGYISLPKSEFVVDLELP
SWLTGNYRIESVLSSSGRKRGICKIAASLKG1"
ORIGIN
Alignment Scores:
Pred. No.: 1,54e-106 Length: 1043
Score: 112.00 Matches: 191
Percent Similarity: 97.95% Conservative: 0
Best Local Similarity: 97.95% Mismatches: 2
Query Match: 58.03% Indels: 4
DB: 9 Gaps: 0
US-10-030-937-9 (1-193) x HUMGM2A (1-1043)
QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db 91 ATGCAGTCCCTGATGAGGCTCCCTCTGATCGCCCTGGCTTCTTCGGGACCCCT 150
QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerPheSerTrpAspAsnCysPhe 40
Db 151 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTTCCTGGGATAACTGTGA- 209
QY 41 -GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa 60
Db 210 TGAAGGGAAGGACCTTCGGTGATCAGAGCCCTGACTTGGAGCCTGACCCCATCGTGT 269
QY 60 lProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLe 80
Db 270 TCCTGGAATGTGACCTCAGTGTCTGGTGGGACGACGAGTGTCCCTGAGTTCTCTCT 329
QY 80 ulysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysTh 100
Db 330 GAAGGTGGATTTAGTTTGTGAGGAAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGCAC 389
QY 100 rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120

Db 390 AGACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTGACATGTAATTC 449
QY 120 oThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPh 140
Db 450 TACTGGGAGCCCTGCCAGAGCCCTGCTGCTACTATGAGCTTCTTGGCCTGCTGCTT 509
QY 140 eLysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuLeuP 160
Db 510 CAAAGAAGGAACCTACTCTACCTGCCAAGACGGAATTCGT-TGTGCTGACCTGGAGCTGC 568
QY 160 roSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysA 180
Db 569 CCAAGTTGGCTACACACGGGAACCTACCGCATAGAGCGCTCTGAGCAGCAGTGGGAAGC 628
QY 180 rGluGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 629 GTCTGGGCTGCATCAAGATCGCTGCCTCTCTAAAGGGCATA 669

RESULT 4
BC009273 2413 bp mRNA linear PRI 29-JUN-2004
LOCUS Homo sapiens GM2 ganglioside activator, mRNA (cDNA clone MGC:10462
DEFINITION IMAGE:4053681), complete cds.
ACCESSION BC009273
VERSION BC009273.2 GI:38197023
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2413)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S.S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, W.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2413)
Strausberg, R.
Direct Submission
Submitted (12-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

TITLE NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Nov 6, 2003 this sequence version replaced gi:14424506.
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,

REMARK
COMMENT

Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Latic,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantripoop,S., Thomas,P.J., Touchman,J.W., Tsurgon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,J.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 15 Row: j Column: 22

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 16507969.

FEATURES

Location/Qualifiers
1..2413
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:10462 IMAGE:4053681"
/tissue_type="Uterus, leiomyosarcoma"
/clone_lib="NIH MGC 46"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

gene

1..2413
/gene="GM2A"
/note="synonym: SAP-3"
/db_xref="LocusID:2760"
/db_xref="MIM:272750"

CDS

30..611
/gene="GM2A"
/codon_start=1
/product="GM2 ganglioside activator, precursor"
/protein_id="AAH09273.1"
/db_xref="GI:14424507"
/db_xref="LocusID:2760"
/db_xref="MIM:272750"
/translation="MQSIQAPLLIALGLLAAQAHLKKPSQLSSFWNCDGKD
FAVRLSLTEFDIVVGNVTLVSVGSPKPLKDLVLEKEVAGLWIKIPCTDY
IGSCTFEFCVDLMDLIPGCPPEPLRTYGLPCHPCPKGTYSLPKSEFVVPDLPL
SWLTTGNRYRISLVSSGKRLGKIKIASLKG1"

ORIGIN

Alignment Scores:
Pred. No.: 3.25e-106 Length: 2413
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservative: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 9 Gaps: 0

US-10-030-937-9 (1-193) x BC009273 (1-2413)

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
DB 150 GNAGGAGAGACCTCGGTGATCAGAGCCTGACTCTGGAGCCTGACCCCATCTCGTT 209
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProIleu 80
DB 210 CCTGAAATGTGACCTCAGTGTCTGTGGGAGCAGCAGTGTCCCTCAGTTCTCTCTG 269
QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrIleLysIleProCysThr 100
DB 270 AAGTGGATTTAGTTTGGAGGAGGAGGTGGCTGGCTCTGGATCAAGATCCATGCACA 329
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
DB 330 GACTACATTGGCAGCTGATACCTTTGAACATCTCTGTGATGCTTGACATGTTAATTCCT 389
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
DB 390 ACTGGGAGCCCTGCCAGAGCCCTGGTACCTATGGGCTTCCTTGGCACTGTCCCTTC 449
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160

DB 450 AAAGAAGAACTACTACTCCCAAGAGCGAATTCGT-TGTGCTGACTGGAGCTGCC 508
QY 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
DB 509 CAGTTGGCTACCAACCGGAACTACCGCATAGAGCGTCTCTGAGCAGCAGTGGGAAGCG 568
QY 180 gLeuGlyCysIleLysIleAlaIleSerLeuLysGlyIle 193
DB 569 TCTGGCTGCATCAAGATCGCTGCTCTCTCTAAAGGGCATA 608
RESULT 5
AX330938 2436 bp DNA linear PAT 09-JAN-2002
LOCUS AX330938
DEFINITION Sequence 1447 from Patent WO0194629.
ACCESSION AX330938
VERSION AX330938.1 GI:18121572
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
JOURNAL gene sets
PATENT: WO 0194629-A 1447 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Location/Qualifiers
1..2436
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 3.27e-106 Length: 2436
Score: 112.00 Matches: 191
Percent Similarity: 97.95% Conservative: 0
Best Local Similarity: 97.95% Mismatches: 2
Query Match: 58.03% Indels: 4
DB: 6 Gaps: 0
US-10-030-937-9 (1-193) x AX330938 (1-2436)
QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
DB 59 ATGCAGTCCCTGATGCAGGCTCCCTCCTGATCGCCCTGGGCTTGTCTTCGCGACCCCT 118
QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspSerCysPhe 40
DB 119 GCGAAGGCCACCTGAAAAGCCATCCAGCTCAGTAGCTTTCTCGGATTAAGTGA- 177
QY 41 -GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa 60
DB 178 TGAAGGAGAGACCTCGGTGATCAGAGCCTGACTCTGGAGCCTGACCCCATCTCGTCT 237
QY 60 lProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProIle 80
DB 238 TCCTGAAATGTGACCTCAGTGTCTGTGGGAGCAGCAGTGTCCCTCAGTTCTCTCT 297
QY 80 uLysValAspLeuValLeuGluLysGluValAlaGlyLeuThrIleLysIleProCysTh 100
DB 298 GAAGTGGATTTAGTTTGGAGAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGCAC 357
QY 100 rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120
DB 358 AGACTACATTGGCAGCTGATACCTTTGAACATCTCTGTGATGCTTGACATGTTAATTC 417
QY 120 oThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPh 140
DB 418 TACTGGGAGCCCTGCCAGAGCCCTGCGTACCTATGGGCTTCCTTGGCACTGTCCCTTC 477


```

FEATURES
  source
    Location/Qualifiers
      1..648
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="pGM2A"
        /tissue_type="placenta"
        /clone_lib="placenta lambda gt11"
        <1..313
          /function="activator of GM2 hydrolysis"
          /codon_start=2
          /product="GM2-activator protein"
          /protein_id="CAA43407.1"
          /db_xref="GI:31855"
          /db_xref="FEMBL:Q14427"
          /translation="KDPVAVRSLEPPDPIVPGNVLTSLVVGSTSVPLSSFLKVDLVL
          KEKVGAGLWIKIPCTDYIGSCFHFCDVLDMLIPTGEPCEPLRTYGLPCHPKFEVS
          T"
      CDS
        3..226-94
          Length: 648
          Score: 100.00
          Matches: 100
          Percent Similarity: 100.00%
          Conservative: 0
          Best Local Similarity: 100.00%
          Mismatches: 0
          Query Match: 51.81%
          Indels: 0
          Gaps: 0
      US-10-030-937-9 (1-193) x HSGM2APA (1-648)
      QY 43 LysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValValProGly 62
      Db 2 AAGGACCTCGGTGATCAGAGCCTGACTCTGGAGCCTGACCCCATGCTGCTCTCGA 61
      QY 63 AsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeuLysVal 82
      Db 62 AATGTGACCTCAGTGTGCTGGCAGCACCAGTGTGCTGCTGCTGCTGCTGCTGCTG 121
      QY 83 AspLeuValLeuGluLysGluValIleGlyLeuThrPheLeuAspMetLeuLeuProThrGly 102
      Db 122 GATTTAGTTTGGAGAGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 181
      QY 103 IleGlySerCysThrPheLeuHisPheCysAspValLeuAspMetLeuLeuProThrGly 122
      Db 182 ATTGGCAGCTGTACCTTTGAACACTCTCTGTGATGCTGTGATGCTGTGATGCTGTG 241
      QY 123 GluProCysProGluProLeuArgThrThrGlyLeuProCysHisCysProPheLysGlu 142
      Db 242 GAGCCCTGCCAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
      RESULT 8
      CQ728078
      LOCUS
      DEFINITION
      ACCESSION
      VERSION
      KEYWORDS
      SOURCE
      ORGANISM
        Homo sapiens (human)
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
      1
      Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
      Kite, such as nucleic acid arrays, comprising a majority of
      humanexons or transcripts, for detecting expression and other uses
      thereof
      Patent: WO 02068579-A 14012 06-SEP-2002;
      PE Corporation (NY) (US)
      JOURNAL
      FEATURES
        source
          1..1045
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
  
```

```

ORIGIN
  Alignment Scores:
  Pred. No.: 2.54e-76 Length: 1045
  Score: 83.00 Matches: 123
  Percent Similarity: 98.40% Conservative: 0
  Best Local Similarity: 98.40% Mismatches: 1
  Query Match: 43.01% Indels: 2
  DB: 6 Gaps: 0
  US-10-030-937-9 (1-193) x CQ728078 (1-1045)
  QY 70 GlySerThrSerValProLeuSerSerProLeuLysValAspLeuValLeuGluLysGlu 89
  Db 299 GGCAGCACCATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 358
  QY 90 ValAlaGlyLeuThrPheLeuLysLeuProCysThrAspTyrIleGlySerCysThrPheGlu 109
  Db 359 GTGGCTGGCTCTGGATCAAGATCCATGTCAGACTACATTGGCAGCTGTACCTTTGAA 418
  QY 110 HisPheCysAspValLeuAspMetLeuLeuProThrGlyGluProCysProGluProLeu 129
  Db 419 CACTTCTGTGATGCTTGCATGTTAATCTCTACTGGGAGCCCTGCCAGAGCCCTG 478
  QY 130 ArgThrTyrGlyLeuProCysHisCysProPheLysGluGlyThrTyrSerLeuProLys 149
  Db 479 CGTACCTATGGCTTCTTGTCCACTGTCTTCAAGAGGAGGAACTACTCTACTGCCAAG 538
  QY 150 SerGluPheAla-ValProAspLeuGluLeuProSerTrpLeuThrThrGlyAsnTyrAr 169
  Db 539 AGCGAATTCGT-TGTGCTGACCTGGAGCTGCCAGTTCCTACCCCGGAGACTACCG 597
  QY 169 GileGluSerValLeuSerSerSerGlyValArgLeuGlyCysIleLysIleAlaLys 189
  Db 598 CATAGAGAGCTCTCTGACGACGAGGAGGAGCGTCTGGCTGCATCAAGATCGTCCCTC 657
  QY 189 rLeuLysGlyIle 193
  Db 658 TCTAAGGGCATA 670
  RESULT 9
  HSGM2APB
  LOCUS
  DEFINITION
  ACCESSION
  VERSION
  KEYWORDS
  SOURCE
  ORGANISM
    HSGM2APB
    H sapiens RNA for GM2-activator protein (clones pGAP2 & pGAP3).
    X61095.1 GI:31856
    G(M2) activator protein.
    Homo sapiens (human)
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (bases 1 to 1093)
    Nagarajan,S., Chen,H.C., Li,S.C., Li,Y.T. and Lockyer,J.M.
    Evidence for two cDNA clones encoding human GM2-activator protein
    Biochem. J. 282 (Pt 3), 807-813 (1992)
    MEDLINE
    PUBMED
    REFERENCE
    2 (bases 1 to 1093)
    Lockyer,J.
    Direct Submission
    Submitted (26-JUL-1991) J. Lockyer, Tulane University Medical
    school, Human Genetics Program, 1430 Tulane Ave., New Orleans LA
    70112, USA
    Location/Qualifiers
      1..1093
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="pGAP2 and pGAP3"
        /tissue_type="placenta"
        /clone_lib="placenta lambda gt11"
        <1..603
          /function="activator of GM2 hydrolysis"
      CDS
  
```



```

/codon_start=1
/product="GM2-activator protein"
/protein_id="CAA43408.1"
/db_xref="GI:31857"
/db_xref="TrEMBL:Q14428"
/translations="RAGPPPMQSLMOAPLIALGILLAAPOAHILKPKSOLSSFSWD
NCDGKDPVIRSLTLEPDLIYPGNVTLVSMGSTVPLSSPLKVLVLEKEVAGNMI
KIFDYGICSTFHEHCDVLDMLIPGEPCEPRITVGLPCHCPFKEGITVSLPSKSFV
VPDLPSWJTTGNRIEYSLSSGKRLGCIKIAASLKG1"
mat_peptide
121..600
/product="unnamed"

ORIGIN
Alignment Scores:
Pred. No.: 2,65e-76 Length: 1093
Score: 83.00 Matches: 123
Percent Similarity: 98.40% Conservatives: 0
Best Local Similarity: 98.40% Mismatches: 1
Query Match: 43.01% Indels: 2
DB: 9 Gaps: 0

US-10-030-937-9 (1-193) x HSGM2APB (1-1093)

QY 70 GlySerThrValProLeuSerSerProLeuLysValAspLeuValLeuGlu 89
Db 229 GGCAGACACAGTGTCCCTCCCTGAGTTCTCTCTGAGGTGGATTAGTTTGGAGAAGGAG 288
QY 90 ValAlaGlyLeuTrpIleLysPheProCysThrAspTyrIleGlySerCysThrPheGlu 109
Db 289 GTGGCTGGCCCTCTGGATCAAGATCCCATGCACAGACTACATTTGGAGTTTGGAA 348
QY 110 HisPheCysAspValLeuAspMetLeuIleProThrGlyGluProCysProGluProLeu 129
Db 349 CACTTCTGTGATGTGCTGTGACATGTTAAATCTCTACTGGGAGGCCCTGCCAGAGCCCTG 408
QY 130 ArgThrTyrGlyLeuProCysHisCysProPheLysGluGlyThrTyrSerLeuProLys 149
Db 409 CGTACCTATGGGCTTCTCTGGCACTGTCCTTCAAGAAGAACCTACTCACTGCCCAAG 468
QY 150 SerGluPheAla-ValProAspLeuGluLeuProSerTrpLeuThrThrGlyAsnTyrAr 169
Db 469 AGCGAAATCGT-TGTGCTGACCTGGAGCTGCCAGTTGGCTCACCACCGGAACTACCG 527
QY 169 gllleGluSerValLeuSerSerGlyLysArgLeuGlyCysIleLysleAlaLase 189
Db 528 CATAGAGAGCGTCTGAGCAGCAGTGGAGCGTCTGGCTGCATCAAGATCGTGCCCTC 587
QY 189 rLeuLysGlyIle 193
Db 588 TCTAAGGGGCATA 600

RESULT 10
CQ050684/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

CQ050684
Sequence 1504 from Patent WO0157270.
CQ050684
CQ050684.1 GI:41025190
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Penn.S.G., Hanzel,D.K., Chen,W. and Rank,D.R.
Human genome-derived single exon nucleic acid probes useful for
analysis of gene expression in human breast and hbl 100 cells
Patent: WO 0157270-A 1504 09-AUG-2001;
Asomica, Inc. (US)
Location/Qualifiers
1..475
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="MAP TO AC011342.1-EXPRESSED IN HELA, SIGNAL = 2.2"

ORIGIN
Alignment Scores:
Pred. No.: 1.06e-53 Length: 475
Score: 61.00 Matches: 61
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.61% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-9 (1-193) x CQ050684 (1-475)

QY 82 ValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThrAsp 101
Db 465 GTGGATTAGTTTGGAGAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGCACAGAC 406
QY 102 TyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIleProThr 121
Db 405 TACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTGACATGTTAAATTCCTACT 346
QY 122 GlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPheLys 141
Db 345 GGGAGCCCTGCCAGAGCCCTCGTACCTATGGGCTTCTTGGCACTGTCCCTTCAAA 286
QY 142 Glu 142
Db 285 GAA 283

RESULT 11
CQ065722/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

CQ065722
Sequence 1522 from Patent WO0157278.
CQ065722
CQ065722.1 GI:41035591
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Penn.S.G., Hanzel,D.K., Chen,W. and Rank,D.R.
Human genome-derived single exon nucleic acid probes useful for
analysis of gene expression in human hela cells or other human
cervical epithelial cells
Patent: WO 0157278-A 1522 09-AUG-2001;
Asomica, Inc. (US)
Location/Qualifiers
1..475
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="MAP TO AC011342.1-EXPRESSED IN HELA, SIGNAL = 2.2"

ORIGIN
Alignment Scores:
Pred. No.: 1.06e-53 Length: 475
Score: 61.00 Matches: 61
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.61% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-9 (1-193) x CQ065722 (1-475)

QY 82 ValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThrAsp 101
Db 465 GTGGATTAGTTTGGAGAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGCACAGAC 406
QY 102 TyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIleProThr 121
Db 405 TACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTGACATGTTAAATTCCTACT 346

```

QY 122 GlyGluProCysProGluProLeuArgThrTyGlyLeuProCysHisCysProPheLys 141
 |||||
 Db 345 GGGGAGCCCTGCCAGAGCCCTGGCTACCTATGGCTTCCTTGCACACTGTCCTTCAAA 286

QY 142 Glu 142
 |||||
 Db 285 GAA 283

RESULT 12
 CQ092707/c
 LOCUS Sequence 1566 from Patent WO0157272.
 DEFINITION
 ACCESSION CQ092707
 VERSION CQ092707.1 GI:41061733
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Penn, S.G., Hanzel, D.K., Chen, W. and Rank, D.R.
 TITLE Human genome-derived single exon nucleic acid probes useful for
 analysis of gene expression in human placenta
 JOURNAL Patent: WO 0157272-A 1566 09-AUG-2001;
 Aeomica, Inc. (US)
 FEATURES Location/Qualifiers
 source 1..475
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="MAP TO AC011342.1-EXPRESSED IN PLACENTA, SIGNAL = 11"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.06e-53 Length: 475
 Score: 61.00 Matches: 61
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.61% Indels: 0
 DB: 6 Gaps: 0

US-10-030-937-9 (1-193) x CQ092707 (1-475)
 QY 82 ValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThrAsp 101
 |||||
 Db 465 GTGGATTAGTTTGGAGAAGGAGTGCTGGCTCTGGATCAAGATCCCATGCACAGAC 406

QY 102 TyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIleProThr 121
 |||||
 Db 405 TACATGGCAGCTGTACCTTTGAACACTTCTGTGATGCTTGACATGTTAATTCCTACT 346

QY 122 GlyGluProCysProGluProLeuArgThrTyGlyLeuProCysHisCysProPheLys 141
 |||||
 Db 345 GGGGAGCCCTGCCAGAGCCCTGGCTACCTATGGCTTCCTTGCACACTGTCCTTCAAA 286

QY 142 Glu 142
 |||||
 Db 285 GAA 283

RESULT 13
 CQ131515/c
 LOCUS Sequence 1537 from Patent WO0157276.
 DEFINITION
 ACCESSION CQ131515
 VERSION CQ131515.1 GI:41088871
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1

AUTHORS Penn, S.G., Hanzel, D.K., Chen, W. and Rank, D.R.
 TITLE Human genome-derived single exon nucleic acid probes useful for
 analysis of gene expression in human bone marrow
 JOURNAL Patent: WO 0157276-A 1537 09-AUG-2001;
 Aeomica, Inc. (US)
 FEATURES Location/Qualifiers
 source 1..475
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="MAP TO AC011342.1-EXPRESSED IN BONE MARROW, SIGNAL = 1.7"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.06e-53 Length: 475
 Score: 61.00 Matches: 61
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.61% Indels: 0
 DB: 6 Gaps: 0

US-10-030-937-9 (1-193) x CQ131515 (1-475)
 QY 82 ValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThrAsp 101
 |||||
 Db 465 GTGGATTAGTTTGGAGAAGGAGTGCTGGCTCTGGATCAAGATCCCATGCACAGAC 406

QY 102 TyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIleProThr 121
 |||||
 Db 405 TACATGGCAGCTGTACCTTTGAACACTTCTGTGATGCTTGACATGTTAATTCCTACT 346

QY 122 GlyGluProCysProGluProLeuArgThrTyGlyLeuProCysHisCysProPheLys 141
 |||||
 Db 345 GGGGAGCCCTGCCAGAGCCCTGGCTACCTATGGCTTCCTTGCACACTGTCCTTCAAA 286

QY 142 Glu 142
 |||||
 Db 285 GAA 283

RESULT 14
 CQ170122/c
 LOCUS Sequence 1518 from Patent WO0157274.
 DEFINITION
 ACCESSION CQ170122
 VERSION CQ170122.1 GI:41164858
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Penn, S.G., Hanzel, D.K., Chen, W. and Rank, D.R.
 TITLE Human genome-derived single exon nucleic acid probes useful for
 analysis of gene expression in human heart
 JOURNAL Patent: WO 0157274-A 1518 09-AUG-2001;
 Aeomica, Inc. (US)
 FEATURES Location/Qualifiers
 source 1..475
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="MAP TO AC011342.1-EXPRESSED IN HEART, SIGNAL = 1.5"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.06e-53 Length: 475
 Score: 61.00 Matches: 61
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.61% Indels: 0
 DB: 6 Gaps: 0

```

US-10-030-937-9 (1-193) x CQ170122 (1-475)

QY      82 ValAspLeuValLeuGluLysGluValAlaGlyLeuTyrIleLysileProCysThrAsp 101
Db      465 GTGGATTAGTTTGGAGAGAGAGTGGCTGGCTCTGGATCAAGATCCCATGACAGAC 406

QY      102 TyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIleProThr 121
Db      405 TACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTGACATGTTAAATTCCTACT 346

QY      122 GlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysPropheLys 141
Db      345 GGGAGCCCTGCCAGAGCCCTCGGTACCTATGGGCTTCCTGGCAGCTGTCCTTCAAA 286

QY      142 Glu 142
Db      285 GAA 283

RESULT 15
CQ199273/c
LOCUS   CQ199273              475 bp      DNA      linear      PAT 21-JAN-2004
DEFINITION
Sequence 1550 from Patent WO0157271.
ACCESSION
CQ199273
VERSION
CQ199273.1 GI:41184937
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Penn,S.G., Hanzel,D.K., Chen,W. and Rank,D.R.
TITLE
Human genome-derived single exon nucleic acid probes useful for
analysis of gene expression in human breast and bt 474 cells
JOURNAL
Patent: WO 0157271-A 1550 09-AUG-2001;
Aeomica, Inc. (US)
FEATURES
Location/Qualifiers
source
1..475
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="MAP TO AC011342.1-EXPRESSED IN BT474, SIGNAL = 1.9"

ORIGIN

Alignment Scores:
Pred. No.:      1.06e-53      Length:      475
Score:          61.00      Matches:      61
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      31.61%      Indels:      0
DB:              6      Gaps:      0

US-10-030-937-9 (1-193) x CQ199273 (1-475)

QY      82 ValAspLeuValLeuGluLysGluValAlaGlyLeuTyrIleLysileProCysThrAsp 101
Db      465 GTGGATTAGTTTGGAGAGAGAGTGGCTGGCTCTGGATCAAGATCCCATGACAGAC 406

QY      102 TyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIleProThr 121
Db      405 TACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTGACATGTTAAATTCCTACT 346

QY      122 GlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysPropheLys 141
Db      345 GGGAGCCCTGCCAGAGCCCTCGGTACCTATGGGCTTCCTGGCAGCTGTCCTTCAAA 286

QY      142 Glu 142
Db      285 GAA 283

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 19, 2004, 00:22:32 ; Search time 510 Seconds
(without alignments)
1986.546 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 193

Sequence: 1 MSLMQAPLLIALGLLALTP.....LSSSGKRLGCIKIAASLKGI 193

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USFTO spool_p/US10030937/runat 16112004 153101 3276/app query.fasta_1.391
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937 @CNG 1.1.885 @runat 16112004 153101 3276 -NCPV=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseq 23Sep04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	58.0	953	12	ADQ17711
2	112	58.0	1043	4	Aaf54705 Human sof
3	112	58.0	1043	4	Aaf54705 Nucleotid
4	112	58.0	1935	10	Adb47402 Nucleotid
5	112	58.0	2384	6	Adb47402 Human cdn
6	112	58.0	2436	3	Abk34915 Human cDN
					Aac55714 Human GM2

Ab163110	Breast ca	6	ABL63110	2436	58.0	112	7
Add71046	Human GM2	10	ADD71046	2436	58.0	112	8
Adn95859	Human BEC	11	ADN95859	2436	58.0	112	9
Aas64907	DNA encod	5	AAS64907	2471	58.0	112	10
Abv78068	Hypoxia-r	6	ABV78068	2478	58.0	112	11
Adn03619	Antipsori	12	ADN03619	2478	58.0	112	12
Aas81113	DNA encod	5	AAS81113	2498	58.0	112	13
Adq22367	Human sof	12	ADQ22367	3988	58.0	112	14
Aas64904	DNA encod	5	AAS64904	546	51.8	100	15
Rai11589	Probe #15	4	AAI11589	475	31.6	61	16
Abas3278	Human foe	4	ABA3278	475	31.6	61	17
Rai32880	Probe #15	4	AAI32880	475	31.6	61	18
Aba42855	Human bre	4	ABA42855	475	31.6	61	19
Aba23052	Probe #15	4	ABA23052	475	31.6	61	20
Rak26980	Human bon	4	AAK26980	475	31.6	61	21
Rak01535	Human bra	4	AAK01535	475	31.6	61	22
ABS26569	Human liv	4	ABS26569	475	31.6	61	23
Rai01513	Probe #15	5	AAI01513	475	31.6	61	24
ABS01567	Human gen	6	ABS01567	475	31.6	61	25
Raf54707	Nucleotid	4	AAF54707	1705	31.6	61	26
Raf54701	Nucleotid	4	AAF54701	1706	31.6	61	27
Raf54704	Nucleotid	4	AAF54704	1706	31.6	61	28
Rai20803	Probe #10	4	AAI20803	145	24.9	48	29
ABA65871	Human foe	4	ABA65871	145	24.9	48	30
Rai46033	Probe #14	4	AAI46033	145	24.9	48	31
Aba47983	Human bon	4	ABA47983	145	24.9	48	32
ABA32957	Probe #11	4	ABA32957	145	24.9	48	33
Rak40013	Human bon	4	AAK40013	145	24.9	48	34
Rak14282	Human bra	4	AAK14282	145	24.9	48	35
ABS39600	Human liv	4	ABS39600	145	24.9	48	36
Rai06511	Probe #65	5	AAI06511	145	24.9	48	37
ABS14099	Human gen	6	ABS14099	145	24.9	48	38
Raf54700	Nucleotid	4	AAF54700	1047	21.8	39	39
Raf54703	Nucleotid	4	AAF54703	1047	21.8	40	40
Raf54706	Nucleotid	4	AAF54706	1047	21.8	41	41
ABQ60520	Human col	6	ABQ60520	448	20.7	42	42
ABZ80229	Human tra	8	ABZ80229	249999	20.7	43	43
Raf54702	Nucleotid	4	AAF54702	633	14.0	27	44
Raf54699	Nucleotid	4	AAF54699	633	14.0	27	45

ALIGNMENTS

RESULT 1
ADQ17711
ID ADQ17711 standard; DNA; 953 BP.
XX
AC ADQ17711;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 528.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression

PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.

PS Example 2; SEQ ID NO 528; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytosolic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.

XX SQ Sequence 953 BP; 201 A; 286 C; 231 G; 235 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,79e-102 Length: 953
 Score: 112.00 Matches: 152
 Percent Similarity: 98.70% Conservative: 0
 Best Local Similarity: 98.70% Mismatches: 1
 Query Match: 58.03% Indels: 2
 DB: 12 Gaps: 0

US-10-030-937-9 (1-193) x AD017711 (1-953)

QY 41 GluGlyAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 DB 211 GAAGGAAAGGACCTCGGTGATCAGAGGCTGACTCTGGAGCCTGACCCCATGCTGTT 270
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 DB 271 CCTGGAATGTGACCTCAGTGTCTGGGAGCAGCAGTGTCTCCCTGAGTTCTCTCTG 330
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
 DB 331 AAGGTGGATTAGTTTGGAGAGAGGTGGCTGCTCTGATCAAGATCCCATGCACA 390
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuPro 120
 DB 391 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTCTTGACATGTTAATTCCT 450
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 DB 451 ACTGGGGAGCCCTGCCCCAGAGCCCTGGGTACCTATGGGCTTCTTGCCACTGTCCCTTC 510
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
 DB 511 AAAGAGAGNACTACTACTGCCAAGAGCGAATTCGT-TGTGCTCTGACCTGGAGCTGCC 569
 QY 160 sSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysAr 180
 DB 570 CAGTTGGCTCACCACGGGAACTACCGCATAGAGAGCGTCTTGAGCAGCAGTGGGAAGCG 629
 QY 180 gIeuGlyCysIleLysIleAlaAserIleLysGlyIle 193
 DB 630 TCTGGGCTGCATCAAGATCGCTGCCTCTCTAAAGGGCATA 669

RESULT 2

ID AAF54705 standard; DNA; 1043 BP.

XX AAF54705;

AC AAF54705;

XX 15-MAY-2001 (first entry)

XX Nucleotide sequence of a human polynucleotide sequence.

DE

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

XX Homo sapiens.

OS WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INNR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX MPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX Claim 11; Page 181-182; 209pp; French.

XX The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX SQ Sequence 1043 BP; 222 A; 303 C; 247 G; 271 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,95e-102 Length: 1043
 Score: 112.00 Matches: 191
 Percent Similarity: 97.9% Conservative: 0
 Best Local Similarity: 97.9% Mismatches: 2
 Query Match: 58.03% Indels: 4
 DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x AAF54705 (1-1043)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuLeuAlaThrPro 20
 DB 91 ATGCAGTCCCTGATGACGCTCCCTCTGATGCGCTGGCTTGTCTTCGCGACCCCT 150
 QY 21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 DB 151 GCGCAAGCCCACTGAAAGCCATCCAGCTCAGTAGCTTTCTTCTGGGATAACTGTGA- 209
 QY 41 -GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa 60
 DB 210 TGAAGGGAGGACCTTGGGTGATCAGAAGCTGACTCTGGAGCTGACCCCATGCTGT 269
 QY 60 IProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProle 80
 DB 270 TCTTGGAAATGTGACCTCAGTGTCTGGGAGCAGCAGTGTCCCTCTGAGTTCTCTCT 329

QY 80 ulysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleIysIleProCysTh 100
 Db 330 GAAGGTGGATTAGTTTGGAGAGAGAGGTGGCTGGCTTGGATCAAGATCCATGCAC 389
 QY 100 rAspTyrlleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120
 Db 390 AGACTACATGGCAGCTGTACCTTGAACACCTCTGTGATGTGCTGACATGTATTTC 449
 QY 120 oThrGlyGluProCysProGluProLeuArgThrTyrlleGlyLeuProCysHisCysProPh 140
 Db 450 TACTTGGGAGCCCTGCCAGAGCCCTGGCTACTATGGCTTCCCTTGGCCTTCCCTT 509
 QY 140 elysGluGlyThrTyrlleSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuP 160
 Db 510 CAAAGAGGAACCTACTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 568
 QY 160 roSerTrpLeuThrThrGlyAsnTyrlleArgIleGluSerValLeuSerSerGlyLysA 180
 Db 569 CCAGTGGCTCACCACCGGGAACCTACCGCATAGAGAGAGAGAGAGAGAGAGAG 628
 QY 180 rGluGlyCysIleIysIleAlaAlaSerLeuLysGlyIle 193
 Db 629 GTCTGGGCTGCATCAAGATGGCTGCTCTCTAAAGGGGCATA 669

RESULT 3

AAF54708
 ID AAF54708 standard; DNA; 1043 BP.
 XX

AAF54708;

15-MAY-2001 (first entry)

Nucleotide sequence of a human polynucleotide sequence.

Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

Homo sapiens.

WO200105422-A2.

25-JAN-2001.

17-JUL-2000; 2000WO-FR002057.

15-JUL-1999; 99FR-00009372.

(INMR) BIOMERIEUX STELHYS.

Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

WPI; 2001-159475/16.

Detecting, preventing and treating degenerative, neurological and
 autoimmune diseases, particularly multiple sclerosis, using specified
 polypeptides or related nucleic acid or ligand.

Claim 11; Page 183; 209pp; French.

The present sequence represents a human polynucleotide sequence, which is
 used in the method of the invention. The specification describes a method
 which uses at least one polypeptide or polynucleotide sequence belonging
 to the perlecan, precursor of the retinol-binding plasma protein.
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid

CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 1043 BP; 222 A; 303 C; 247 G; 271 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.95e-102 Length: 1043
 Score: 112.00 Matches: 191
 Percent Similarity: 97.95% Conservative: 0
 Best Local Similarity: 97.95% Mismatches: 2
 Query Match: 58.03% Indels: 4
 DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x AAF54708 (1-1043)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuAlaThrPro 20
 Db 91 ATGCAGTCCCTGATGCAGGCTCCCTTCCATGCGCCCTGGGCTTGGTTCTCGGACCCCT 150
 QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 Db 151 GCGCAAGCCACCTGAAAGAGCCATCCAGCTCAGTAGCTTTCTCTGGGATACTGTGA- 209
 QY 41 -GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValva 60
 Db 210 TGAAGGGAG 269
 QY 60 lProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProle 80
 Db 270 TCTTGGAAATGTGACCTCAGTGTGTGTGGGAGACACAGTGTCCCTGAGTTCTCTCT 329
 QY 80 ulysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysTh 100
 Db 330 GAAGGTGGATTAGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 389
 QY 100 rAspTyrlleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120
 Db 390 AGACTACATGGCAGCTGTACCTTGAACACCTCTGTGATGTGCTGACATGTATTTC 449
 QY 120 oThrGlyGluProCysProGluProLeuArgThrTyrlleGlyLeuProCysHisCysProPh 140
 Db 450 TACTTGGGAGCCCTGCCAGAGCCCTGGCTACTATGGCTTCCCTTGGCCTTCCCTT 509
 QY 140 elysGluGlyThrTyrlleSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuP 160
 Db 510 CAAAGAGGAACCTACTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 568
 QY 160 roSerTrpLeuThrThrGlyAsnTyrlleArgIleGluSerValLeuSerSerGlyLysA 180
 Db 569 CCAGTGGCTCACCACCGGGAACCTACCGCATAGAGAGAGAGAGAGAGAGAGAG 628
 QY 180 rGluGlyCysIleIysIleAlaAlaSerLeuLysGlyIle 193
 Db 629 GTCTGGGCTGCATCAAGATGGCTGCTCTCTAAAGGGGCATA 669

RESULT 4

ADB47402
 ID ADB47402 standard; cDNA; 1935 BP.
 XX
 AC ADB47402;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Human cDNA upregulated in dendritic cells SEQ ID NO 102.
 XX
 KW ss; gene; human; dendritic cells; high throughput; cancer;
 KW infectious disease; autoimmune disease; allergy;
 KW graft versus host disease; vaccine enhancing; gene therapy.
 XX
 OS Homo sapiens.
 XX


```

Query Match: 58.03% Indels: 2
DB: 6 Gaps: 0
US-10-030-937-9 (1-193) x ABK34915 (1-2384)
QY 41 GluGlyLysAspProAlaValIleAsgSerLeuThrLeuGluProAspProIleValVal 60
D 133 GAAGGGAAGACCTCGGTGATCAGAGCCTGACTCTGGAGCTGACCCATCGTCGTT 192
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
D 193 CCTGGAAATGTGACCTCAGTGTGCTGGGAGCAGCAGTGTCCCTGAGTTCTCTCTG 252
QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrIleLysLeuProCysThr 100
D 253 AAGGTGATTTAGTTTGGAGAGGAGTGGCTGGGCTCTGGATCAAGATCCCATGCACA 312
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
D 313 GACTACATTTGGGAGCTGTACCTTTGAACACTTCTGTGATGTGCTTGACATGTTAATTC 372
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
D 373 ACTGGGAGCCCTGCCAGAGCCCTCGGTACCTATGGGCTTCTTGCCACTGTCCCTTC 432
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
D 433 AAAGAGAGGAACCTACTCACTGCCCAAGAGCGAATTCGT-TGTGCTTGACCTGGAGCTGC 491
QY 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysAr 180
D 492 CAGTTGGCTCACACCGGGAATACCCGATAGAGAGCGTCTGAGCAGCAGTGGGAGCG 551
QY 180 gluGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
D 552 TCTGGGCTGCATCAGATCGTCTCTCTAAAGGGCATA 591

RESULT 6
ID AAC55714 standard; cDNA; 2436 BP.
XX
AC AAC55714;
XX
DT 17-JAN-2001 (first entry)
XX
DE Human GM2 activator protein cDNA sequence from Genbank X62078.
XX
KW Human; differentially regulated gene; macrophage development; diagnosis;
KW matrix metalloproteinase 19; MMP19; antiarthritic; antiinflammatory;
KW destructive macrophage development inhibitor; arthritis;
KW colorectal cancer; immune response; ss.
XX
OS Homo sapiens.
XX
EN WO200055373-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-US006883.
XX
PR 15-MAR-1999; 99US-0124530P.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI Murray R;
XX
DR WPI; 2000-628200/60.
XX
PT Screening drug candidates comprises adding a drug to a cell expressing an
PT expression profile gene and determining the effect of the drug on the
PT expression of the expression profile gene.
XX
PS Claim 1; Page; 99pp; English.

```

```

XX The present invention describes a method for screening drug candidates.
XX The method comprises adding a drug to a cell that expresses an expression
XX profile gene encoding a protein encoded by 5 sequences of defined base
XX pairs as given in C55638, C55642, C55643, C55644 and C55653 or a sequence
XX represented by Genbank accession number X92521, X62466, J04130, X62087
XX and X76534 (or a fragment) and determining the effect of the drug on the
XX expression of the expression profile gene. An inhibitor of matrix
XX metalloproteinase 19 (MMP-19), preferably an antibody, is useful for
XX treating destructive macrophage disorders (DMD) by inhibiting DM
XX development in a cell of an individual having arthritis. Antibodies to
XX MMP-19 are useful for localising a therapeutic moiety preferably
XX cytotoxic agent or a radioisotope to colorectal cancer tissue. A
XX composition comprising MMP-19 is useful for eliciting an immune response
XX in an individual. C55635 to C55710 represent human differentially
XX regulated genes of the invention. The present human sequence represents the
XX human GM2 activator protein cDNA sequence according to the Genbank
XX accession number X62078. N.B. The present sequence is not given in the
XX present specification, but it is specifically claimed by its Genbank
XX accession number
XX
SQ Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 4,38e-102 Length: 2436
Score: 112.00 Matches: 191
Percent Similarity: 97.95% Conservative: 0
Best Local Similarity: 97.95% Mismatches: 2
Query Match: 58.03% Indels: 4
DB: 3 Gaps: 0
US-10-030-937-9 (1-193) x AAC55714 (1-2436)
QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuAlaThrPro 20
D 59 ATGCAGTCCCTGATGCAGGCTCCCTCTCTGATCGCCCTGGGCTTGGTTCTCGGACCCCT 118
QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
D 119 GCGCAAGCCACCTGAAAAGGCATCCAGCTCAGTAGCTTTTCTGGGATAACTGTGA- 177
QY 41 -GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa 60
D 178 TGAAGGGAAGGACCCCTGCGGTGATCAGAAGCCCTGACTCTGGAGCCCTGACCCCATCGTC 237
QY 60 lProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLe 80
D 238 TCTTGGAAATGTGACCTCAGTGTGCTGGGAGCACCAGTGTCCCTCCTGAGTTCTCTCT 297
QY 80 uLysValAspLeuValLeuGluLysGluValAlaGlyLeuThrIleLysIleProCysTh 100
D 298 GAAGGTGGATTTAGTTTGGAGAAGGAGGTGGCTGGCTCTGGATCAGATCCATGCCAC 357
QY 100 rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120
D 358 AGACTACATTTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTTGACATGTAATTC 417
QY 120 oThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPh 140
D 418 TACTGGGAGCCCTGCCAGAGCCCTGCTACCTATGGGCTTCTTGGCCATGTCCTCT 477
QY 140 eLysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuP 160
D 478 CAAAGAAGGAACCTACTCACTGCCCAAGAGCGAATTCGT-TGTGCTGACCTGGAGCTGC 536
QY 160 roSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysA 180
D 537 CCAGTTGGCTCACCCCGGAACTACCGCATAGAGAGCGTCTGAGCAGCAGTGGGAAGC 596
QY 180 rgLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
D 597 GTCTGGGCTGCATCAAGATCGCTGCTCTCTAAAGGGCATA 637

```

RESULT 7
ABL63110
ID ABL63110 standard; DNA; 2436 BP.
XX AC ABL63110;
XX XX
XX 15-MAY-2002 (first entry)
XX
XX Breast cancer related gene sequence SEQ ID NO:1447.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US010838.
XX
XX 05-JUN-2000; 2000US-0209473P.
XX 05-JUN-2000; 2000US-0209531P.
XX 18-SEP-2000; 2000US-0233113P.
XX 18-SEP-2000; 2000US-0233113P.
XX 20-SEP-2000; 2000US-0234009P.
XX 20-SEP-2000; 2000US-0234034P.
XX 20-SEP-2000; 2000US-0234052P.
XX 22-SEP-2000; 2000US-0234509P.
XX 22-SEP-2000; 2000US-0234567P.
XX 25-SEP-2000; 2000US-0234923P.
XX 25-SEP-2000; 2000US-0234924P.
XX 25-SEP-2000; 2000US-0235077P.
XX 25-SEP-2000; 2000US-0235082P.
XX 25-SEP-2000; 2000US-0235134P.
XX 25-SEP-2000; 2000US-0235280P.
XX 26-SEP-2000; 2000US-0235637P.
XX 26-SEP-2000; 2000US-0235638P.
XX 27-SEP-2000; 2000US-0235711P.
XX 27-SEP-2000; 2000US-0235720P.
XX 27-SEP-2000; 2000US-0235840P.
XX 27-SEP-2000; 2000US-0235863P.
XX 28-SEP-2000; 2000US-0236028P.
XX 28-SEP-2000; 2000US-0236032P.
XX 28-SEP-2000; 2000US-0236033P.
XX 28-SEP-2000; 2000US-0236034P.
XX 28-SEP-2000; 2000US-0236109P.
XX 29-SEP-2000; 2000US-0236111P.
XX 29-SEP-2000; 2000US-0236842P.
XX 29-SEP-2000; 2000US-0236891P.
XX 02-OCT-2000; 2000US-0237116P.
XX 02-OCT-2000; 2000US-0237172P.
XX 02-OCT-2000; 2000US-0237173P.
XX 02-OCT-2000; 2000US-0237278P.
XX 02-OCT-2000; 2000US-0237294P.
XX 02-OCT-2000; 2000US-0237295P.
XX 02-OCT-2000; 2000US-0237316P.
XX 03-OCT-2000; 2000US-0237425P.
XX 03-OCT-2000; 2000US-0237598P.
XX 03-OCT-2000; 2000US-0237604P.
XX 03-OCT-2000; 2000US-0237606P.
XX 03-OCT-2000; 2000US-0237608P.
XX 01-NOV-2000; 2000US-0244867P.
XX 01-NOV-2000; 2000US-0245084P.
XX
XX (AVALON) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX

PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
XX in expression of a gene of a signature gene set.
XX
XX Claim 1; SEQ ID NO 1447; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
XX tumour
XX
XX Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 4.38e-102 Length: 2436
Score: 112.00 Matches: 191
Percent Similarity: 97.95% Conservative: 0
Best Local Similarity: 97.95% Mismatches: 2
Query Match: 58.03% Indels: 4
DB: Gaps: 0
US-10-030-937-9 (1-193) x ABL63110 (1-2436)
QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
DB 59 ATGCAGTCCCTGATGCAGGCTCCCTCTGATGCCCTGGGCTTGCTTCGCGACCCCT 118
QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerThrAspAsnCysPhe 40
DB 119 GCGCAAGGCCACCTGMAAAGCCATCCAGCTCAGTAGCTTTCTCTGGGTAACCTGTGA- 177
QY 41 -GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa 60
DB 178 TGAAGGGGAAGGACCCCTGCGGTGATCAGAAAGCTGACTCTGGAGCTGACCCCATCGTCT 237
QY 60 lProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProle 80
DB 238 TCCTGGAAATGTGACCTCAGTGTGCGGAGACACAGTGTCCCTGAGTTCTCTCT 297
QY 80 uLysValAspLeuValLeuGluLysGluValAlaGlyLeuThrIleLysIleProCysTh 100
DB 298 GAAGGTGGATTAGTTTGGAGAAAGAGGTGGCTGCTGGATCAAGATCCCATGCAC 357
QY 100 rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120
DB 358 AGACTACATTTGGCAGCTGTACTTTGAACACTTCTGTGATGCTGATGTTAATTC 417
QY 120 oThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPh 140
DB 418 TACTGGGAGCCCTGCCAGAGCCCTGCGTACCTATGGGCTTCCTGCCACTGTCCTT 477
QY 140 eLysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuP 160
DB 478 CAAGAGAGAACTACTACTCTGCCAGAGAGGAATTCTGT-TGTGCTGACCTGGAGCTGC 536
QY 160 roSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysA 180
DB 537 CCAGTTGGCTCACCAACCGGGAACCTACCCCATAGAGAGGCTCTGAGCAGCAGTGGAGC 596
QY 180 rgLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193

597 GTCTGGCTGCATCAAGATCGCTGCTCTCTAAAGGGCATA 637

DB	DB:	10	Gaps:	0
US-10-030-937-9 (1-193) x ADD71046 (1-2436)				
RESULT 8				
ADD71046				
ID	ADD71046 standard; DNA; 2436 BP.			
XX				
AC	ADD71046;			
XX				
DT	15-JAN-2004 (first entry)			
XX				
DE	Human GM2 ganglioside activated protein gene SEQ ID NO:50.			
XX				
KW	liver cancer; chronic hepatitis; cirrhosis; liver disease; hepatotropic;			
KW	cytostatic; gene therapy; human; gene; ds.			
XX				
OS	Homo sapiens.			
XX				
PN	WO2003061564-A2.			
XX				
PD	31-JUL-2003.			
XX				
PF	20-DEC-2002; 2002WO-US040718.			
XX				
PR	21-DEC-2001; 2001US-0341815P.			
PR	31-DEC-2001; 2001US-0343185P.			
XX				
PA	(GENE-) GENE LOGIC INC.			
PA	(LGBI-) LG BIOMEDICAL INST.			
XX				
PI	Koh SS, Liu Q, Chung H, Zeng W, Lee B, Yeramilli S, Song SY;			
XX				
DR	WPI; 2003-663343/62.			
XX				
PT	Diagnosing liver cancer cells, useful for treating liver cancer			
PT	associated with chronic hepatitis or cirrhosis comprises detecting the			
PT	level of expression in a tissue sample of one or more genes associated			
PT	with cancerous liver tissues.			
XX				
PS	Claim 1; SEQ ID NO 50; 176pp; English.			
XX				
CC	The present invention describes a method for diagnosing liver cancer			
CC	cells comprising detecting the level of expression in a tissue sample of			
CC	one or more genes given in the specification (see ADD70997 to ADD71105),			
CC	where differential expression of the genes is indicative of liver cancer.			
CC	Also described: (1) detecting the progression of liver cancer in a			
CC	patient; (2) monitoring the treatment of a patient with liver cancer; (3)			
CC	treating a patient with liver cancer; (4) typing a liver disease in a			
CC	patient; (5) detecting the presence or progression of liver cancer in a			
CC	patient with chronic hepatitis or cirrhosis; (6) differentiating liver			
CC	cancer related to chronic hepatitis from liver cancer related to			
CC	cirrhosis; (7) screening for an agent capable of modulating the onset or			
CC	progression of liver cancer; (8) a composition comprising at least two			
CC	oligonucleotides comprising a sequence that specifically hybridizes to			
CC	any of the genes; (9) a solid support comprising the at least two			
CC	oligonucleotides; (10) a computer system comprising a database containing			
CC	information identifying the level in liver tissue of a set of genes; (11)			
CC	a method for using the computer system to present information identifying			
CC	the expression level in tissue or cell of any of the genes; and (12) a			
CC	therapeutic agent for slowing or halting the progression of liver cancer.			
CC	The methods are useful for treating liver cancer associated with chronic			
CC	hepatitis or cirrhosis. The present sequence represents a specifically			
CC	claimed human gene sequence which is used in the exemplification of the			
CC	present invention.			
XX				
SQ	Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;			
Alignment Scores:				
Pred. No.:	4,388-102	Length:	2436	
Score:	112.00	Matches:	191	
Percent Similarity:	97.95%	Conservative:	0	
Best Local Similarity:	97.95%	Mismatches:	2	
Query Match:	58.03%	Indels:	4	

QY	1	MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuLeuAlaThrPro	20
DB	59	ATGCAGTCCCTGATGAGGCTCCCTCCTGATCGCCCTGGGCTTGTCTTCGGAGCCCT	118
QY	21	AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe	40
DB	119	GGCAAGCCACCTGAAAAAGCCATCCAGCTCAGTAGCTTTTCTCTGGGATACTGTGA	177
QY	41	-GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa	60
DB	178	TGAAGGGAGAGCCCTGCGGTGATCAGAAGCCCTGACTCTGGAGCCTGACCCCATCGT	237
QY	60	lProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLe	80
DB	238	TCTTGGAAATGTGACCTCAGTGTCTGGGAGCACCAGTGTCCCTTGTGAGTTCCTCT	297
QY	80	uLysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysTh	100
DB	298	GAAGGTGGATTTAGTTTTTGAGAAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGC	357
QY	100	rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr	120
DB	358	AGACTACATTTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGTGACATGTAATTCC	417
QY	120	oThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPh	140
DB	418	TACTGGGGAGCCCTGCCAGACCCCTGCGTACCTATGGGCTTCTTGGCACGTGCCCT	477
QY	140	eLysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuP	160
DB	478	CAAAGAAGGAACCTACTCTACTGCCCAAGAGCGAATTTCGT-TGTGCTGTACCTGGAGCTGC	536
QY	160	rSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysA	180
DB	537	CCAGTTGGCTCACCACCGGAACTACCGCATAGAGAGCGTCTCTGAGCAGCAGTGGGAAGC	596
QY	180	rgLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle	193
DB	597	GTCTGGGCTGCATCAAGATCGCTCTCTTAAAGGGCATA	637
RESULT 9			
ADN95859			
ID	ADN95859 standard; DNA; 2436 BP.		
XX			
AC	ADN95859;		
XX			
DT	01-JUL-2004 (first entry)		
XX			
DE	Human BEC/LEC-related gene sequence SeqID783.		
XX			
KW	growth; differentiation; blood endothelial cell; BEC;		
KW	lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;		
KW	lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;		
KW	vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;		
KW	inflammatory disease; cancer metastasis; lymphatic system; gene; ds;		
XX	human.		
OS	Homo sapiens.		
XX			
PN	WO2003080640-A1.		
XX			
PD	02-OCT-2003.		
XX			
PF	07-MAR-2003; 2003WO-US006900.		
XX			
PR	07-MAR-2002; 2002US-0363019P.		
XX			
PA	(LUDW-) LUDWIG INST CANCER RES.		
PA	(LICN) LICENTIA LTD.		

XX PI Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
 DR WPI; 2003-876899/81.
 DR P-PSDB; ADN95858.
 XX Example 1; SEQ ID NO 783; 176pp; English.
 XX This invention relates to a method of differentially modulating the
 CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
 CC endothelial cells (LEC) comprises contacting endothelial cells with a
 CC composition comprising an agent that differentially modulates blood or
 CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
 CC identifying a human subject with lymphoedema and with a mutation in at
 CC least one allele of a gene encoding a LEC protein, where the mutation
 CC correlates with lymphoedema in human subjects, and with the proviso that
 CC the LEC protein is not VEGFR-3; and administering to the subject a
 CC composition comprising a lymphatic growth agent selected from VEGF-C or
 CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
 CC the development of compounds with an antiangiogenic, cytostatic,
 CC vasotropic or antiinflammatory activity or for gene therapy. The method
 CC is useful in modulating the growth or differentiation of blood
 CC endothelial cells or lymphatic endothelial cells, in treating hereditary
 CC lymphoedema, in screening for an endothelial cell disorder or
 CC predisposition to the disorder or in monitoring the efficacy or toxicity
 CC of a drug on endothelial cells. The agent is useful in manufacturing a
 CC medicament for the differential modulation of blood vessel endothelial
 CC cell or lymphatic vessel endothelial cell growth or differentiation. The
 CC lymphatic growth agent may also be used in manufacturing a medicament for
 CC the treatment of hereditary lymphoedema resulting from a mutation in a
 CC LEC gene or of other diseases involving the lymphatic vessels, such as
 CC various inflammatory diseases and cancer metastasis via the lymphatic
 CC system. The present sequence is that of a human LEC/BEC differentially
 CC expressed gene which is related to the method of the invention. Note: This
 CC sequence does not appear in the specification but was obtained by the
 CC indexer using the source data given in table 14 of the specification.
 XX SQ Sequence 2436 BP; 570 A; 632 C; 576 G; 558 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4:38e-102 Length: 2436
 Score: 112.00 Matches: 191
 Percent Similarity: 97.95% Conservative: 0
 Best Local Similarity: 97.95% Mismatches: 2
 Query Match: 58.03% Indels: 4
 DB: 11 Gaps: 0
 US-10-030-937-9 (1-193) x ADN95859 (1-2436)
 QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 DB 59 ATGCAGTCCCTGATCAGGCTCCCTCTGATCGCCCTGGGCTTCTTCTGCGACCCCT 118
 QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTirAspAsnCysPhe 40
 DB 119 GCGCAGGCCACCTGAAAAGCCATCCAGCTAGCTTTCTTCTGGGTAACCTGTA- 177
 QY 41 -GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa 60
 DB 178 TGAAGGGAAGGACCCCTGGCGTATCAGAAGCCCTGACTCTGGAGCTGACCCCATCGTGT 237
 QY 60 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProle 80
 DB 238 TCTTGGAATGACCCCTGAGTGTGTGGGAGACACCTAGTGTCCCTCTGAGTTCCTCT 297
 QY 80 uLysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPrlIleLysIleProCysTh 100
 DB 298 GAAGGTGATTTAGTTTGGAGAAGGAGGCTGGCTGCTGATCAAGATCCCATGCAC 357
 QY 100 rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120
 DB 358 AGACTACATGGCAGCTGTACCTTTGAACACTTCTGTGATGTGTGATGCTTAATTC 417

QY 120 oThrGlyGluProCysProGluProLeuAlaGlyThrTyrGlyLeuProCysHisCysProPh 140
 DB 418 TACTGGGGAGCCCTGCCCAGAGCCCTTGGTACCTATGGGCTTCTTGGCACTGTCCTT 477
 QY 140 eLysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuP 160
 DB 478 CARAAGAGGAACCTACTACTCTGCCCAAGAGCGAATTGCT-TGTGCTTGACCTGGAGCTGC 536
 QY 160 roSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysA 180
 DB 537 CCAGTTGGCTCACCCAGCGGAACCTACCGCATAGAGCGCTCTGAGCAGCAGTGGGAAGC 596
 QY 180 rGLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 DB 597 GTCTGGGCTGATCAAGATCGTCTGCTCTCTAAAGGGCATA 637
 RESULT 10
 AAS64907
 ID AAS64907 standard; cDNA; 2471 BP.
 AC AAS64907;
 XX 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #711.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US008631.
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 DR Drmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG00720.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 PS Claim 1; SEQ ID NO 711; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 2471 BP; 578 A; 639 C; 583 G; 671 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4,466-102 Length: 2471
 Score: 112.00 Matches: 152
 Percent Similarity: 98.70% Conservative: 0
 Best Local Similarity: 98.70% Mismatches: 1
 Query Match: 58.03% Indels: 2
 DB: 5 Gaps: 0

US-10-030-937-9 (1-193) x AA564907 (1-2471)

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 213 GAAGGGAAGGACCTCGCGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTGT 272
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 Db 273 CTTGAAATGTGACCTCAGTGTCTGGGACGACGAGTGTCCCTGAGTTCTCTCTCG 332
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
 Db 333 AAGTGGATTTAGTTTGGAGAGGAGTGGCTGGCTCTGGATCAAGATCCCATGCACA 392
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 393 GACTACATTTGGCAGCTGTACCTTTGAACACTCTGTGATGTCTTGACATGTTAATTCCT 452
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 453 ACTGGGAGCCCTGGCCAGAGCCCTCGTACCTATGGCTTCTTGGCCACTGTCCCTTC 512
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPr 160
 Db 513 AAAGAGGAGACCTACTCACTGCCAAGACGCAATTCGT-TGTGCCCTGACCTGGAGCTGCC 571
 QY 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
 Db 572 CAGTTGGCTCACCCCGGAACTACCGCATAGAGAGCGTCTCTGAGCAGCAGTGGGAGCG 631
 QY 180 GluGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db 632 TCTGGGCTGCATCAAGATCGTGCCTCTCTAAAGGGCATA 671

RESULT 11

ABV78068

ID ABV78068 standard; DNA; 2478 BP.

XX AC ABV78068;

XX DT 12-NOV-2002 (first entry)

XX DE Hypoxia-regulated protein coding sequence #88.

XX KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;
 XX KW antiinflammatory; vulnary; gynecological; ophthalmological; vaccine;
 XX KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
 XX KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;
 XX KW preclapsmia; atherosclerosis; inflammatory condition; wound healing;
 XX KW inflammation; erythropoiesis; hair loss; human; gene; ds.

XX OS Homo sapiens.

XX PN WO200246465-A2.

XX PD 13-JUN-2002.

XX PF 10-DEC-2001; 2001WO-GB005459.

XX PD 08-DEC-2000; 2000GB-00030076.

PR 08-FEB-2001; 2001GB-00003156.

PR 25-OCT-2001; 2001GB-00025666.

XX (OXFO-) OXFORD BIOMEDICA UK LTD.

XX White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;
 PI Rayner WN;

XX WPI; 2002-627238/67.

PT Identifying a gene involved in disease for treating hypoxia-regulated
 PT conditions, comprises comparing the transcriptome/proteome of two cell
 PT types under different conditions and identifying a differentially
 PT regulated gene.

XX Claim 37; Page 397-398; 538pp; English.

CC The present invention relates to methods for identifying genes and
 CC proteins that are implicated in a specific disease or physiological
 CC condition. The method comprises comparing the transcriptome/proteome of a
 CC specialised cell type implicated in a disease or condition with that of a
 CC second specialised cell type, under two experimental conditions, and
 CC identifying a gene that is differentially regulated in the two
 CC specialised cell types under experimental conditions. ABV7873-ABV78116
 CC and ABV65061-ABV65257 were identified using the methods of the invention.
 CC The coding sequences and proteins are useful for treating a disease in a
 CC patient, for manufacture of a medicament for treating hypoxia-regulated
 CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
 CC biological response to hypoxia conditions, or hypoxic-associated
 CC pathology in a patient. The coding sequences and proteins are also useful
 CC for monitoring the therapeutic treatment of a disease or physiological
 CC condition, such as cancer, ischaemic conditions, reperfusion injury,
 CC retinopathy, neonatal stress, preclapsmia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss

XX SQ Sequence 2478 BP; 588 A; 638 C; 586 G; 666 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4,466-102 Length: 2478
 Score: 112.00 Matches: 191
 Percent Similarity: 97.95% Conservative: 0
 Best Local Similarity: 97.95% Mismatches: 2
 Query Match: 58.03% Indels: 4
 DB: 6 Gaps: 0

US-10-030-937-9 (1-193) x ABV78068 (1-2478)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuLeuIlePro 20
 Db 96 ATGCAGTCCCTGATCGAGCTCCCTCCTGATCGCCCTGGGCTTCTTCGCGACCCCT 155
 QY 21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerThrAspAsnCysPhe 40
 Db 156 GCGCAAGCCACCTGAAAAGCCATCCAGCTCAGTAGCTTTCCTGGATCACTGTGA- 214
 QY 41 -GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa 60
 Db 215 TGAAGGGAAGGACCCCTGGCGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTGT 274
 QY 60 lProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProle 80
 Db 275 TCCTGGAATGTGACCCCTCAGTGTCTGGGAGCAGCACCAGTGTCCCCCTGAGTTCCTCT 334
 QY 80 uLysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysTh 100
 Db 335 GAAGGTGATTTAGTTTGGAGAGAGAGTGGCTGGCTCTGGATCAAGATCCCATGAC 394
 QY 100 rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120
 Db 395 AGACTACATTGGCAGCTGTACCTTTGAACACTCTCTGATGTGCTTGACATGTTAATTC 454
 QY 120 oThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPh 140

Db 455 TACTGGGAGCCCTGCCAGAGCCCTGGCTGCTATGGGCTTCCTGCCACTGTCCTT 514

Qy 140 eLysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuP 160

Db 515 CAAGAAGGAACCTACTCACTGCCCAAGAGCGAATTGCT-TGTGCTTGACCTGGAGCTGC 573

Qy 160 roSerTrpLeuThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysA 180

Db 574 CCAGTTGGCTCACACCGGGAACCTACCGCATAGAGAGCGTCTCGAGCAGCAGTGGGAAGC 633

Qy 180 rgLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193

Db 634 GTCTGGGCTGCATCAAGATCGCTGCTCTCTAAAGGGGCATA 674

RESULT 12

ADN03619

ID ADN03619 standard; cDNA; 2478 BP.

XX AC ADN03619;

DT 01-JUL-2004 (first entry)

XX DE Antipsoriatic cDNA sequence #7.

XX KW ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.

XX OS Homo sapiens.

FN WO2004028479-A2.

XX PD 08-APR-2004.

XX PF 25-SEP-2003; 2003WO-US030907.

XX PR 25-SEP-2002; 2002US-0414006P.

XX PA (GETH) GENENTECH INC.

XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;

XX FI Wu TD;

XX DR WPI; 2004-305105/28.

XX DR P-PSDB; ADN03620.

XX PT New PRO nucleic acid or polypeptide, useful for preparing a

PT pharmaceutical composition for diagnosing or treating psoriasis in a

XX mammal.

PS Claim 1; SEQ ID NO 13; 3069pp; English.

XX CC The invention relates to novel polynucleotide and polypeptides for

CC treating psoriasis or a sequence having at least 80% identity to the

CC above sequences. The nucleic acid is useful for preparing a composition

CC for diagnosing or treating psoriasis in a mammal. This sequence

CC corresponds to one of the polynucleotides of the invention.

XX SQ Sequence 2478 BP; 588 A; 638 C; 586 G; 666 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.46e-102	Length:	2478
Score:	112.00	Matches:	191
Percent Similarity:	97.95%	Conservative:	0
Best Local Similarity:	97.95%	Mismatches:	2
Query Match:	58.03%	Indels:	4
DB:	12	Gaps:	0

US-10-030-937-9 (1-193) x ADN03619 (1-2478)

Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuAlaThrPro 20

Db 96 ATGCAGTCCCTGATCAGGCTCCCTCTGTCGCTGGGCTTCTTCGCGACCCCT 155

Qy 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40

Db 156 GCGCAAGCCACCTGAAAGAGCCATCCAGCTCAGTAGCTTTTCTGGGTAACGTGA- 214

Qy 41 -GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa 60

Db 215 TGAAGGGAAGGACCCCTCGGCTGATCAGAAGCCTGACTCTGGAGCTGACCCCATCGTCT 274

Qy 60 lProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLe 80

Db 275 TCCTGGAAATGTGACCCCTCAGTGTGCTGGGAGCACCAGTGTCCCTGAGTTCTCCCTCT 334

Qy 80 uLysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysTh 100

Db 335 GAAGTGGATTAGTTTGGAGAGGAGTGGCTGGCTCTGGATCAAGATCCCATGCAC 394

Qy 100 rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120

Db 395 AGACTACATTGGCAGCTGTACTTTGAACACTTCTGTGATGTGCTTGACATGTTAAATTC 454

Qy 120 oThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPh 140

Db 455 TACTGGGAGCCCTGCCAGAGAGCCCTGCGTACCTATGGGCTTCTTGCCACTGTCCCTT 514

Qy 140 eLysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuP 160

Db 515 CAAGAAGGAACCTACTCACTGCCCAAGAGCGAATTGCT-TGTGCTTGACCTGGAGCTGC 573

Qy 160 roSerTrpLeuThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysA 180

Db 574 CCAGTTGGCTCACACCGGGAACCTACCGCATAGAGAGCGTCTCGAGCAGCAGTGGGAAGC 633

Qy 180 rgLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193

Db 634 GTCTGGGCTGCATCAAGATCGCTGCTCTCTAAAGGGGCATA 674

RESULT 13

AAS81113/c

ID AAS81113 standard; cDNA; 2498 BP.

XX AC AAS81113;

DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #16917.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; AEG16926.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

XX biodiversity.

PS Claim 1; SEQ ID NO 16917; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 2498 BP; 669 A; 593 C; 646 G; 589 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 4,49e-102 Length: 2498
Score: 112.00 Matches: 191
Percent Similarity: 97.95% Conservative: 0
Best Local Similarity: 97.95% Mismatches: 2
Query Match: 58.03% Indels: 4
DB: 5 Gaps: 0

US-10-030-937-9 (1-193) x AAS81113 (1-2498)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
DB 2440 ATGCAGTCCCTGTATGCAGGCTCCCTCTCATGCTGGCTTGTCTTCGGAGCCCT 2381
QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerThrAspAsnCysPhe 40
DB 2380 GCGCAGAGCCACCTGAAAGAGCCATCCAGTCAGTAGCTTTCTCTGGGATACTGTGA- 2322
QY 41 -GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa 60
DB 2321 TGAAGGAGAGAGCCCTCGGTGTATCAGAACCTTGACTCTGGAGCCTGACCCCATGCTGCT 2262
QY 60 lProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLe 80
DB 2261 TCCTGGAAATGTGACCTCAGTGTCTGTGGGAGCAGCAGTGTCCCTGAGTTCCTCTCT 2202
QY 80 uLysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysTh 100
DB 2201 GAAGGTGGATTAGTTTGGAGAGAGAGGTGGTGGCTCTGGATCAAGATCCATGCAC 2142
QY 100 rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120
DB 2141 AGACTACATATGGCAGTGTACCTTTGACACTCTCTGTGATGCTGTGACATGATATTC 2082
QY 120 oThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPh 140
DB 2081 TACTGGGGAGCCCTGCGCAGAGCCCTGCGTACCTATGGGCTTCTTCCACATGTCCTT 2022
QY 140 eLysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuP 160
DB 2021 CAAAGAGAGGAACCTACTCATCTGCCAAGAGCGAATTCGT-TGTGCTGTGACCTGGAGCTGC 1963
QY 160 roSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysA 180
DB 1962 CCAGTTGGCTCACCACCGGAACCTACCGCATAGAGAGCGTCCCTGAGCAGCGAGGGAAGC 1903
QY 180 rgLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
DB 1902 GTCTGGGCTGCATCAAGATCGCTCTCTCTAAAGGGGCATA 1862

RESULT 14

ADQ22367
ID ADQ22367 standard; DNA; 3988 BP.
XX
AC ADQ22367;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5187.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX ds.
XX
XX Homo sapiens.
XX OS
XX WO2004048938-A2.
XX FN
XX 10-JUN-2004.
XX PD
XX 26-NOV-2003; 2003WO-US038193.
XX PF
XX 26-NOV-2002; 2002US-0429739P.
XX PR
XX (PROT-) PROTEIN DESIGN LABS INC.
XX PA
XX Aziz N, Ginsburg WM, Zlotnik A;
XX PI
XX WPI; 2004-441208/41.
XX DR
XX
XX Early detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX and comparing the gene expression, also useful in treating soft tissue
XX sarcoma.
XX
XX Example 2; SEQ ID NO 5187; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the
XX expression of the gene in both soft tissue samples, where a higher level
XX of protein expression in the first soft tissue sample indicates the
XX presence of soft tissue sarcoma. The method of the invention has
XX cytostatic applications and may be useful for detecting soft tissue
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX acid sequences may be useful in diagnostic and screening applications.
XX The current sequence is that of a human soft tissue sarcoma-upregulated
XX DNA of the invention. The current sequence is not shown within the
XX specification per se but was submitted in CD format by the inventor.
XX
XX Sequence 3988 BP; 1094 A; 927 C; 935 G; 986 T; 0 U; 46 Other;

Alignment Scores:

Pred. No.: 7,03e-102 Length: 3988
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservative: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 12 Gaps: 0

US-10-030-937-9 (1-193) x ADQ22367 (1-3988)

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
DB 216 GAAGGAGAGAGCCCTCGGTGTATCAGAGCCTGACTCTGGAGCCTGACCCCATCGCTGT 275
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
DB 276 CTGTGAAATGTGACCTCAGTGTCTGTGGGAGCAGCAGTGTCCCTGTGATCTCTCTG 335
QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100

Db 336 AAGTGGATTAGTTTGGAGAGAGGTGCTGGCTCTGGATCAGATCCCATGCACA 395
 Qy 101 AsptyrileGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 396 GACTACATTGGCAGCTGTACCTTTGAACACATCTCTGTATGTCTTGACATGTTAATTCCT 455
 Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 456 ACTGGGAGCCCTGCCCCAGAGCCCTGCTGCTACCTATGGGCTTCCCTGCCACTGTCCCTTC 515
 Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
 Db 516 AAAGAAGGAACCTACTCTGCTCCCAAGAGCGAATTCGT-TGTGCTGTACCTGGAGCTGCC 574
 Qy 160 oSerTrpLeuThrThrGlyAsnTyrArgileGluSerValLeuSerSerGlyLysAr 180
 Db 575 CAGTTGGCTCACCACCGGAACTACCGCATAGAGAGCGTCTTGAGCAGCAGTGGGAAGCG 634
 Qy 180 gluGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db 635 TCTGGCTGCATCAAGATCGCTGCTCTCTAAAGGGCATA 674
 RESULT 15
 AAS64904
 ID AAS64904 standard; cDNA; 546 BP.
 AC AAS64904;
 XX
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #708.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX
 PN WO2001/5067-A2.
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI; 2001-639362/73.
 DR P-PSDB; ABG00717.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID NO 708; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 546 BP; 123 A; 134 C; 141 G; 148 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.16e-90 Length: 546
 Score: 100.00 Matches: 100
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 51.81% Indels: 0
 DB: 5 Gaps: 0
 US-10-030-937-9 (1-193) x AAS64904 (1-546)
 Qy 43 LysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValValProGly 62
 Db 2 AAGGACCCCTGGCGTATCAGAGCCCTGACTCTGGAGCCCTGACCCCATCGTCTCTGGA 61
 Qy 63 AsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeuLysVal 82
 Db 62 AATGTGACCCCTCAGTGTGTGGGACAGCAGTGTCCCCCTGAGTTCCTCTGAAGGTG 121
 Qy 83 AspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThrAspTyr 102
 Db 122 GATTTAGTTTGGAGAGGAGGTGGCTGCTGGATCAGATCCCATGCAGACTAC 181
 Qy 103 IleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIleProThrGly 122
 Db 182 ATTGGCAGCTGTACCTTTGAACACTTCTGTGTGATGCTTGACATGTTAATCTCTCTGG 241
 Qy 123 GluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPheLysGlu 142
 Db 242 GAGCCCTGCCAGAGCCCTGCTGATCTATGGGCTTCTTGGCACTGTCCCTTCAAGAA 301

Search completed: November 19, 2004, 03:38:22
 Job time : 516 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 19, 2004, 03:12:52 ; Search time 92 Seconds
(without alignments)
1491.111 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 193

Sequence: 1 MSLMQAPLIIALLGLLATP.....LSSGRKLCIKIAASLKGI 193

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10030937/runat_16112004_153102_3316/app.query.fasta_1.391
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=olip2n.rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937 @cgn_1_1_128 @runat_16112004_153102_3316 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	4.7	787	3	US-08-961-083-97
2	9	4.7	787	4	US-09-536-784-97
3	9	4.7	811	3	US-08-961-083-205
4	9	4.7	811	4	US-09-536-784-205
5	9	4.7	876	4	US-09-583-110-1540
6	9	4.7	12127	4	US-08-961-527-148
7	8	4.1	462	4	US-09-252-991A-546
8	8	4.1	858	4	US-09-252-991A-3981
9	8	4.1	1314	4	US-09-252-991A-7168
10	8	4.1	1335	4	US-09-252-991A-3925
11	8	4.1	1395	4	US-09-711-164-270
12	8	4.1	1395	4	US-09-492-709A-121

c 13	8	4.1	1456	3	US-09-308-406-1	Sequence 1, Appli
c 14	8	4.1	1458	4	US-09-252-991A-570	Sequence 570, App
c 15	8	4.1	1482	4	US-09-252-991A-4050	Sequence 4050, Ap
c 16	8	4.1	1605	4	US-09-252-991A-3897	Sequence 3897, Ap
c 17	8	4.1	1677	4	US-09-252-991A-7476	Sequence 7476, Ap
c 18	8	4.1	1692	4	US-09-252-991A-7219	Sequence 7219, Ap
c 19	8	4.1	2223	4	US-09-252-991A-4015	Sequence 4015, Ap
c 20	8	4.1	2401	3	US-09-320-878-20	Sequence 20, Appli
c 21	8	4.1	2401	4	US-09-141-908-9	Sequence 9, Appli
c 22	8	4.1	2401	4	US-09-657-440-20	Sequence 20, Appli
c 23	8	4.1	2430	3	US-09-105-537-23	Sequence 23, Appli
c 24	8	4.1	2772	3	US-08-936-135-1	Sequence 1, Appli
c 25	8	4.1	2772	4	US-09-439-711C-1	Sequence 1, Appli
c 26	8	4.1	2881	2	US-08-570-227A-1	Sequence 1, Appli
c 27	8	4.1	2881	3	US-09-077-991-1	Sequence 1, Appli
c 28	8	4.1	3241	3	US-09-434-288-11	Sequence 11, Appli
c 29	8	4.1	5035	2	US-08-616-392C-3	Sequence 3, Appli
c 30	8	4.1	5653	4	US-09-583-638-1	Sequence 1, Appli
c 31	8	4.1	6774	4	US-09-674-460-1	Sequence 1, Appli
c 32	8	4.1	13613	3	US-09-105-537-3	Sequence 3, Appli
c 33	8	4.1	36741	3	US-09-301-665-3	Sequence 3, Appli
c 34	8	4.1	46718	4	US-09-816-093-3	Sequence 3, Appli
c 35	8	4.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 36	8	4.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c 37	7	3.6	27	3	US-08-985-162-1452	Sequence 1452, Ap
c 38	7	3.6	27	3	US-08-584-040-472	Sequence 472, App
c 39	7	3.6	27	3	US-08-584-040-3423	Sequence 3423, Ap
c 40	7	3.6	27	3	US-08-584-040-6526	Sequence 6526, Ap
c 41	7	3.6	27	3	US-08-584-040-6640	Sequence 6640, Ap
c 42	7	3.6	27	4	US-09-401-063-1452	Sequence 1452, Ap
c 43	7	3.6	168	4	US-09-270-767-2637	Sequence 2637, Ap
c 44	7	3.6	168	4	US-09-270-767-17919	Sequence 17919, A
c 45	7	3.6	216	4	US-09-583-110-2496	Sequence 2496, Ap

ALIGNMENTS

RESULT 1
US-08-961-083-97
; Sequence 97, Application US/08961083
; Patent No. 6159469

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,083

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 97:

; SEQUENCE CHARACTERISTICS:

```
; LENGTH: 787 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-97

Alignment Scores:
Pred. No.: 4.57 Length: 787
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.66% Indels: 0
DB: 3 Gaps: 0

US-10-030-937-9 (1-193) x US-08-961-083-97 (1-787)

QY 66 LeuSerValValGlySerThrSerVal 74
|||||
DB 464 TTGCTGTGTAGGTTCCACTTCAGTA 490

RESULT 2
US-09-536-784-97
; Sequence 97, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 97:
; LENGTH: 787 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 97:

US-09-536-784-97

Alignment Scores:
Pred. No.: 4.57 Length: 787
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.66% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x US-09-536-784-97 (1-787)

; LENGTH: 787 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-97

Alignment Scores:
Pred. No.: 4.57 Length: 787
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.66% Indels: 0
DB: 3 Gaps: 0

US-10-030-937-9 (1-193) x US-08-961-083-97 (1-787)

QY 66 LeuSerValValGlySerThrSerVal 74
|||||
DB 464 TTGCTGTGTAGGTTCCACTTCAGTA 490

RESULT 3
US-08-961-083-205
; Sequence 205, Application US/08961083
; Patent No. 6159489
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 205:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 811 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-083-205

Alignment Scores:
Pred. No.: 4.7 Length: 811
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.66% Indels: 0
DB: 3 Gaps: 0

US-10-030-937-9 (1-193) x US-08-961-083-205 (1-811)

QY 66 LeuSerValValGlySerThrSerVal 74
|||||
DB 488 TTGCTGTGTAGGTTCCACTTCAGTA 514

RESULT 4
US-09-536-784-205
; Sequence 205, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
```

;
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; OPERATING: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 205:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 811 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 205:
US-09-536-784-205

Alignment Scores:
Pred. No.: 4.7 Length: 811
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.66% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x US-09-536-784-205 (1-811)

QY 66 LeuSerValValGlySerThrSerVal 74
|||
DB 488 TTGCTGTTGTTAGGTTCCACTTCAGTA 514

RESULT 5

US-09-583-110-1540
; Sequence 1540, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 1540
; LENGTH: 876
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-1540

Alignment Scores:
Pred. No.: 5.07 Length: 876
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 4.66% Indels: 0
DB: 4 Gaps: 0
US-10-030-937-9 (1-193) x US-09-583-110-1540 (1-876)
QY 66 LeuSerValValGlySerThrSerVal 74
|||
DB 550 TTGCTGTTGTTAGGTTCCACTTCAGTA 576
RESULT 6
US-08-961-527-148
; Sequence 148, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; OPERATING: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-9504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12127 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-148

Alignment Scores:
Pred. No.: 65.9 Length: 12127
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.66% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x US-08-961-527-148 (1-12127)

QY 66 LeuSerValValGlySerThrSerVal 74
|||
DB 624 TTGCTGTTGTTAGGTTCCACTTCAGTA 650

RESULT 7

US-09-252-991A-546/c
; Sequence 546, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 546
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-546

Alignment Scores:
Pred. No.: 29.5 Length: 462
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.15% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x US-09-252-991A-546 (1-462)

QY 71 SerThrSerValProLeuSerSer 78
|||||
DB 395 TCTAGTCAGTCCGCTTCGTCGA 372

RESULT 8

US-09-252-991A-3981/c
; Sequence 3981, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3981
; LENGTH: 858
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (656)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-3981

Alignment Scores:
Pred. No.: 54 Length: 858
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.15% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x US-09-252-991A-3981 (1-858)

QY 11 IleAlaLeuGlyLeuLeuAla 18
|||||
DB 265 ATCGCCCTCGGCTGCTCTTGTCT 242

RESULT 9

US-09-252-991A-7168
; Sequence 7168, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7168
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7168

Alignment Scores:
Pred. No.: 81.8 Length: 1314
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.15% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x US-09-252-991A-7168 (1-1314)

QY 14 GlyLeuLeuAlaThrProAla 21
|||||
DB 89 GGGCTGCTCCTGGCCACGCGGCC 112

RESULT 10

US-09-252-991A-3925
; Sequence 3925, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3925
; LENGTH: 1335
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (93)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-3925

Alignment Scores:
Pred. No.: 83.1 Length: 1335
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.15% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x US-09-252-991A-3925 (1-1335)

QY 11 IleAlaLeuGlyLeuLeuAla 18
|||||
DB 484 ATCGCCCTCGGCTGCTCTTGTCT 507

RESULT 11

US-09-711-164-270/c

Fri Nov 19 14:12:50 2004

us-10-030-937-9.olip2n.rni

QY 61 ProGlyAsnValThrLeuSerVal 68
DB 460 CCTGGAATGTAAACGCTATCAGTT 437

RESULT 13

US-09-308-406-1/c
; Sequence 1, Application US/09308406
; Patent No. 6159696
; GENERAL INFORMATION:
; APPLICANT: Dijkema, Rein
; APPLICANT: van den Wijngaard, Arthur
; TITLE OF INVENTION: Method and materials for the screening of therapeutic
; TITLE OF INVENTION: agents for the prevention and/or treatment of
; TITLE OF INVENTION: osteoporosis
; FILE REFERENCE: 1/96227 US
; CURRENT APPLICATION NUMBER: US/09/308,406
; CURRENT FILING DATE: 1999-06-21
; EARLIER APPLICATION NUMBER: PCT/EP97/06668
; EARLIER FILING DATE: 1997-11-20
; EARLIER APPLICATION NUMBER: EP96203283.5
; EARLIER FILING DATE: 1996-11-22
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1456
; TYPE: DNA
; ORGANISM: human
US-09-308-406-1

Alignment Scores:
Pred. No.: 90.4 Length: 1456
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.15% Indels: 0
DB: 3 Gaps: 0

US-10-030-937-9 (1-193) x US-09-308-406-1 (1-1456)

QY 175 SerSerSerGlyLysArgLeuGly 182
DB 943 TCACGCTCTGGAGCGGCTGGGG 920

RESULT 14

US-09-252-991A-570
; Sequence 570, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 570
; LENGTH: 1458
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-570

Alignment Scores:
Pred. No.: 90.5 Length: 1458
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.15% Indels: 0
DB: 4 Gaps: 0

; Sequence 270, Application US/09711164

; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 270
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1395)
US-09-711-164-270

Alignment Scores:
Pred. No.: 86.7 Length: 1395
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.15% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x US-09-711-164-270 (1-1395)

QY 61 ProGlyAsnValThrLeuSerVal 68
DB 460 CCTGGAATGTAAACGCTATCAGTT 437

RESULT 12

US-09-492-709A-121/c
; Sequence 121, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 121
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: E. Coli
US-09-492-709A-121

Alignment Scores:
Pred. No.: 86.7 Length: 1395
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.15% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x US-09-492-709A-121 (1-1395)

US-10-030-937-9 (1-193) x US-09-252-991A-570 (1-1458)

QY 71 SerThrSerValProLeuSerSer 78
 Db 1364 TCTACGTCAGTTCGCTTCGICA 1387

RESULT 15

US-09-252-991A-4050/c
 ; Sequence 4050, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 4050
 ; LENGTH: 1482
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1065)
 ; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
 US-09-252-991A-4050

Alignment Scores:

Pred. No.:	92	Length:	1482
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.15%	Indels:	0
DB:	4	Gaps:	0

US-10-030-937-9 (1-193) x US-09-252-991A-4050 (1-1482)

QY 11 IleAlaLeuGlyLeuLeuLeuAla 18
 Db 674 ATGCCCTCGGCTGCTCTTGCT 651

Search completed: November 19, 2004, 05:53:00
 Job time : 105 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 19, 2004, 04:55:34 ; Search time 4517 Seconds
(without alignments)
230.883 Million cell updates/sec

Title: US-10-030-937-9
Perfect score: 193
Sequence: 1 MQSLMQAPLIALGLLILATP.....LSSSGKRLGCIKIAASLKGI 193

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3627888 seqs, 2701811610 residues

Word size: 1

Total number of hits satisfying chosen parameters: 7246336

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10030937/runat 16112004 153104 3403/app query.fasta_1.391
-DB=Published Applications NA -OFMT=fastap -SUFFIX=olip2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10030937 @CGN 1 1 912 @runat 16112004 153104 3403
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

Result No.	Score	Query Match	Length	DB ID	Description
1	112	58.0	1935	10	US-09-971-392-102
2	112	58.0	2384	9	US-09-822-849A-53
3	112	58.0	2436	9	US-09-954-531-380
4	112	58.0	2436	10	US-09-525-978B-81
5	112	58.0	2478	15	US-10-170-385-390
c 6	61	31.6	475	9	US-09-864-761-1518
c 7	48	24.9	145	9	US-09-864-761-18277
c 8	40	20.7	448	11	US-09-969-034-4215
c 9	40	20.7	250000	15	US-10-325-610-26
10	18	9.3	546	13	US-10-027-632-207798
11	18	9.3	546	13	US-10-027-632-207799
12	18	9.3	546	13	US-10-027-632-207800
13	18	9.3	546	13	US-10-027-632-207801
14	18	9.3	546	15	US-10-027-632-207798
15	18	9.3	546	15	US-10-027-632-207799
16	18	9.3	546	15	US-10-027-632-207800
17	18	9.3	546	15	US-10-027-632-207801
18	18	9.3	1983	16	US-10-388-934-167
c 20	9	4.7	218	11	US-09-922-293-2287
c 21	9	4.7	455	13	US-10-027-632-202246
c 22	9	4.7	455	15	US-10-027-632-202246
23	9	4.7	649	16	US-10-425-114-7785
24	9	4.7	787	9	US-09-765-272-97
25	9	4.7	811	9	US-09-765-272-205
26	9	4.7	1051	16	US-10-424-599-38496
27	9	4.7	12127	8	US-08-961-527-148
28	8	4.1	60	10	US-09-908-975-22643
c 29	8	4.1	195	9	US-09-864-761-18267
c 30	8	4.1	223	16	US-10-424-599-102858
c 31	8	4.1	274	18	US-10-425-115-5227
c 32	8	4.1	319	9	US-09-864-761-20513
c 33	8	4.1	331	18	US-10-425-115-120690
34	8	4.1	391	11	US-09-864-408A-4241
35	8	4.1	406	9	US-09-960-352-14056
36	8	4.1	420	16	US-10-424-599-85374
c 37	8	4.1	430	9	US-09-864-761-1508
38	8	4.1	458	9	US-09-917-800A-144
39	8	4.1	458	10	US-09-918-995-8914
c 40	8	4.1	467	17	US-10-437-963-18713
c 41	8	4.1	471	9	US-09-917-800A-910
42	8	4.1	484	10	US-09-918-995-30887
c 43	8	4.1	495	10	US-09-918-995-25173
c 44	8	4.1	593	13	US-10-027-632-277778
c 45	8	4.1	593	15	US-10-027-632-277778

ALIGNMENTS

RESULT 1
US-09-971-392-102
; Sequence 102, Application US/09971392
; Publication No. US20030134283A1
; GENERAL INFORMATION:
; APPLICANT: Peterson, David P.
; APPLICANT: Pearson, Cecelia I.
; APPLICANT: Cocks, Benjamin G.
; TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
; FILE REFERENCE: PA-0029 US
; CURRENT APPLICATION NUMBER: US/09/971,392
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 60/237,652
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PERL Program
; SEQ ID NO 102
; LENGTH: 1935
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

NAME/KEY: misc feature
OTHER INFORMATION: Template ID: 977615.8
US-09-971-392-102

Alignment Scores:
Pred. No.: 1,72e-102 Length: 1935
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservativeness: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 10 Gaps: 0

US-10-030-937-9 (1-193) x US-09-971-392-102 (1-1935)

QY 41 GluGlyAspProAlaValIleArgSerLeuThrLeuGluProAspProlleValVal 60
Db 222 GAAGGGAAGACCCCTGCGGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTCTG 281
QY 61 ProGlyAsnValThrLeuSerValGlySerThrSerValProLeuSerSerProLeu 80
Db 282 CCTGGAAATGATGACCTCAGTGTGCTGGGAGCAGCAGTGTCCCTGAGTTCTCTCTG 341
QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
Db 342 AAGGTGATTTAGTTTGGAGAAGAGGTGGCTGGCTCTGGATCAAGATCCCATGCACA 401
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 402 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTTGAATTCCT 461
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 462 ACTGGGAGCCCTGCCAGAGCCCTGCGTACCTATGGGCTTCTTCCACTGTCCTTC 521
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
Db 522 AAAGAAGGAACCTACTCAGTCCCAAGAGCGAATTGCT-TGTGCTGACCTGGAGCTGCC 580
QY 160 oSerTrpLeuThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
Db 581 CAGTTGGCTCACCCCGGAACTACCGCATAGAGAGCGTCTCAGCAGCAGTGGGAAGCG 640
QY 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 641 TCTGGGCTGCATCAAGATGCTGCTCTCTTAAAGGGCATA 680

RESULT 2

US-09-822-849A-53
Sequence 53, Application US/09822849A
Patent No. US20020045170A1
GENERAL INFORMATION:
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakar
APPLICANT: Graham, James R.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6403
CURRENT APPLICATION NUMBER: US/09/822,849A
PRIOR FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/195,582
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 598
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 53
LENGTH: 2384
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-849A-53

Alignment Scores:

Pred. No.: 2,08e-102 Length: 2384
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservativeness: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 9 Gaps: 0

US-10-030-937-9 (1-193) x US-09-822-849A-53 (1-2384)

QY 41 GluGlyAspProAlaValIleArgSerLeuThrLeuGluProAspProlleValVal 60
Db 133 GAAGGGAAGACCCCTGCGGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTCTG 192
QY 61 ProGlyAsnValThrLeuSerValGlySerThrSerValProLeuSerSerProLeu 80
Db 193 CCTGGAAATGATGACCTCAGTGTGCTGGGAGCAGCAGTGTCCCTGAGTTCTCTCTG 252
QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
Db 253 AAGGTGATTTAGTTTGGAGAAGAGGTGGCTGGCTCTGGATCAAGATCCCATGCACA 312
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 313 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTTGAATTCCT 372
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 373 ACTGGGAGCCCTGCCAGAGCCCTGCGTACCTATGGGCTTCTTCCACTGTCCTTC 432
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
Db 433 AAAGAAGGAACCTACTCAGTCCCAAGAGCGAATTGCT-TGTGCTGACCTGGAGCTGCC 491
QY 160 oSerTrpLeuThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
Db 492 CAGTTGGCTCACCCCGGAACTACCGCATAGAGAGCGTCTCAGCAGCAGTGGGAAGCG 551
QY 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 552 TCTGGGCTGCATCAAGATGCTGCTCTCTTAAAGGGCATA 591

RESULT 3

US-09-954-531-380
Sequence 380, Application US/09954531
Patent No. US20020165180A1
GENERAL INFORMATION:
APPLICANT: Weaver, Zoe
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
FILE REFERENCE: 689290-77
CURRENT APPLICATION NUMBER: US/09/954,531
PRIOR FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: US/60/233,133
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,009
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,034
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,509
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US/60/234,567
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 1392
SOFTWARE: Patent In version 3.0
SEQ ID NO 380
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-531-380

Alignment Scores:


```

> PRIOR APPLICATION NUMBER: US 60/180,312
> PRIOR FILING DATE: 2000-02-04
> PRIOR APPLICATION NUMBER: US 60/207,456
> PRIOR FILING DATE: 2000-05-26
> PRIOR APPLICATION NUMBER: US 09/632,366
> PRIOR FILING DATE: 2000-08-03
> PRIOR APPLICATION NUMBER: GB 24263.6
> PRIOR FILING DATE: 2000-10-04
> PRIOR APPLICATION NUMBER: US 60/236,359
> PRIOR FILING DATE: 2000-09-27
> PRIOR APPLICATION NUMBER: PCT/US01/00666
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00667
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00664
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00669
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00665
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00668
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00663
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00662
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00661
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00670
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: US 60/234,687
> PRIOR FILING DATE: 2000-09-21
> PRIOR APPLICATION NUMBER: US 09/608,408
> PRIOR FILING DATE: 2000-06-30
> PRIOR APPLICATION NUMBER: US 09/774,203
> PRIOR FILING DATE: 2001-01-29
> NUMBER OF SEQ ID NOS: 49117
> SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
> SEQ ID NO 1518
> LENGTH: 475
> TYPE: DNA
> ORGANISM: Homo sapiens
> FEATURE:
> OTHER INFORMATION: MAP TO AC011342.1
> OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
> OTHER INFORMATION: EXPRESSED IN HL100, SIGNAL = 2.3
> OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
> OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
> OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
> OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
> OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
> OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
> OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
> OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99
> US-09-864-761-1518

```

Alignment Scores:	
Pred. No.:	8,94e-52
Score:	61.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	31.61%
DB:	9
Length:	475
Matches:	61
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

Score:	61.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	31.61%	Indels:	0
DB:	9	Gaps:	0
US-10-030-937-9 (1-193) x US-09-864-761-1518 (1-475)			
Qy	82	VallaspLeuValLeuGluLysGluValalaGlyLeuTrpIleLysIleProCysThrAsp	101
Db	465	GTGGATTAGTTTGGAGAGAGGCTGGCTCTCGATCAAGATCCCATGCACAGAC	406
Qy	102	TyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeulleProThr	121
Db	405	TACATTGGCAGTCTACCTTTTGACACTCTCTGATGTCCTGATCATGTATTCTTACT	346

RESULT 11
US-10-027

```
; Sequence 207799, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 207799
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-10-027-632-207799

Alignment Scores:
Pred. No.: 1.69e-08 Length: 546
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.33% Indels: 0
DB: 13 Gaps: 0

US-10-030-937-9 (1-193) x US-10-027-632-207799 (1-546)
Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIle 58
Db 297 GAAGGGAAGGACCCCTGCGGTGATCAGAGGCTGACTCTGGAGCCTGACCCCATC 350

RESULT 12
US-10-027-632-207800
; Sequence 207800, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 207800
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-10-027-632-207801

Alignment Scores:
Pred. No.: 1.69e-08 Length: 546
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.33% Indels: 0
DB: 13 Gaps: 0

US-10-030-937-9 (1-193) x US-10-027-632-207801 (1-546)
Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIle 58
Db 297 GAAGGGAAGGACCCCTGCGGTGATCAGAGGCTGACTCTGGAGCCTGACCCCATC 350

RESULT 14
US-10-027-632-207798
; Sequence 207798, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```

; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207798
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-207798

```

```

Alignment Scores:
Pred. No.: 1.69e-08 Length: 546
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.33% Indels: 0
DB: 15 Gaps: 0

```

US-10-030-937-9 (1-193) x US-10-027-632-207798 (1-546)

```

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProfile 58
DB 297 GAAGGAGAGACCCCTCGGGTGATCAGAGCCTGACTCTGGAGCCTGACCCCATC 350

```

RESULT 15

```

US-10-027-632-207799
; Sequence 207799, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207799
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-207799

```

```

Alignment Scores:
Pred. No.: 1.69e-08 Length: 546
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.33% Indels: 0
DB: 15 Gaps: 0
US-10-030-937-9 (1-193) x US-10-027-632-207799 (1-546)
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProfile 58
DB 297 GAAGGAGAGACCCCTCGGGTGATCAGAGCCTGACTCTGGAGCCTGACCCCATC 350

```

Search completed: November 19, 2004, 08:20:42
Job time : 4550 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 19, 2004, 03:02:02 ; Search time 3320 Seconds
(without alignments)
2118.334 Million cell updates/sec

Title: US-10-030-937-9
Perfect score: 193
Sequence: 1 MQSLWQAPLIALGILLATP.....LSSSGKRLCCIKIASLXGI 193

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 1821986598 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlp

-Q=/cgn2_1/USPTO_spool_p/US10030937/runat_16112004_153102_3298/app_query.fasta_1.391
-DB=EST -QFMT=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937@cgn_1_1_6425@runat_16112004_153102_3298 -NCPU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_ges1:*
9: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	153	79.3	784	4	BG323734 602421833
2	135	69.9	885	6	CA453926 AGENCOURT
3	112	58.0	559	1	AU280628 AU280628
4	112	58.0	698	2	BF509172 UI-H-B14-
5	112	58.0	711	4	BG479322 602526351
6	112	58.0	733	4	BI856212 603382936
7	112	58.0	784	4	BI838554 603086219
8	112	58.0	842	4	BI091220 602856051
9	112	58.0	858	4	BI820051 603037236

10	112	58.0	895	5	BQ676659
11	112	58.0	910	1	AL560604
12	112	58.0	937	5	BUI51364
13	112	58.0	948	5	BUS56606
14	112	58.0	949	5	BQ060062
15	112	58.0	973	4	BM474816
16	112	58.0	1018	2	BE735010
17	112	58.0	1031	4	BM476220
18	105	54.4	909	5	BQ643369
19	96	49.7	818	1	AL552056
20	93	48.2	651	4	BM723945
21	93	48.2	912	1	AL543858
22	93	48.2	994	1	AL548441
23	93	48.2	1060	1	AL550565
24	83	43.0	600	5	EX473154
25	83	43.0	601	5	EX506263
26	83	43.0	726	4	BG762599
27	83	43.0	760	4	BG770447
28	83	43.0	847	5	BQ220522
29	83	43.0	863	4	BG478588
30	83	43.0	997	4	BM561693
31	83	43.0	2338	3	CR626644
32	82	42.5	533	1	AI793102
33	80	41.5	465	6	CD710451
34	79	40.9	1033	5	BQ662907
35	75	38.9	476	4	BM147068
36	73	37.8	366	2	BE182886
37	70	36.3	813	4	BG913328
38	68	35.2	1130	2	BE613752
39	67	34.7	438	2	BE041669
40	67	34.7	611	4	BG623044
41	67	34.7	660	4	BG470369
42	66	34.2	712	6	CA422926
43	66	34.2	1067	2	BF528447
44	65	33.7	434	6	CD710184
45	62	32.1	389	2	BF949518

ALIGNMENTS

RESULT 1
BG323734
LOCUS BG323734
DEFINITION 602421833F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4559888 5', mRNA linear EST 27-FEB-2001
ACCESSION BG323734 GI:13130171
VERSION BG323734.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1. (bases 1 to 784)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
High quality sequence stop: 777.
Location/Qualifiers
1. 784
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4559888"

10876659 AGENCOURT
AL560604 AL560604
BUI51364 AGENCOURT
BUS56606 AGENCOURT
BQ060062 AGENCOURT
BM474816 AGENCOURT
BE735010 601567832
BM476220 AGENCOURT
BQ643369 AGENCOURT
AL552056 AL552056
BM723945 UI-E-BO1-
AL543858 AL543858
AL548441 AL548441
AL550565 AL550565
EX473154 DKFP686H
EX506263 DKFP686F
BG762599 602734472
BG770447 602734356
BQ220522 AGENCOURT
BG478588 602524087
BM561693 AGENCOURT
CR626644 full-length
AI793102 on4c06.y
CD710451 EST28978
BQ662907 AGENCOURT
BM147068 TCAAP1Q10
BE182886 CM4-HT065
BG913328 602812047
BE613752 601504554
BE041669 ho33604.x
BG623044 602647926
BG470369 602533946
CA422926 UI-H-FLO-
BF528447 602043611
CD710184 EST26711
BF949518 MR3-NN021

/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 14"
/note="Organ: kidney; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:
Pred. No.: 1.64e-137 Length: 784
Score: 153.00 Matches: 153
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 79.27% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BG323734 (1-784)

Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 126 GAAGGGAAGGACCTCGCTGATCAGAGCTGACTCTGGAGCTGACCCCATGCTGTT 185
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 186 CCTGGAATGTGACCCCTCAGTGTGCTGGCAGCAGCAGTGTCCCTGAGTCTCTCTG 245
Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPheLysIleProCysThr 100
Db 246 AAGGTGGATTAGTTTGGAGAGAGGAGTGGCTGCTGATCAAGATCCCATGCACA 305
Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 306 GACTACATTTGACCTGTACCTTTGAACACTCTGTGATGCTGATGATTTAATCT 365
Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 366 ACTGGGAGCCCTGCCAGAGCCCTGCTACCTATGGCTTCTTGGCACTGTCCTTC 425
Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
Db 426 AAAGAAGGAACCTACTCACTCCCAAGAGCGAATTCGCTGTGCTGACCTGAGCTGCC 485
Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
Db 486 AGTTGGCTCACACCGGGAACCTACCGCATAGAGAGCTCTCTGAGCAGCAGTGGGAAG 545
Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 546 CTGGGCTGCATCAAGATCGCTGCCTCTCTAAAGGCGATA 584

RESULT 2
CA453926 885 bp mRNA linear EST 12-NOV-2002
LOCUS AGENCOURT_10738491 MAPcL Homo sapiens cDNA clone IMAGE:6718573 5',
DEFINITION mRNA sequence.
ACCESSION CA453926
VERSION CA453926.1 GI:24903163
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 885)
NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Kristi A. Eglund, Ira Pastan
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUAM14275 row: b column: 13
High quality sequence stop: 427.
Location/Qualifiers
source
1. .885
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6718573"
/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
hTERT-HME1, LNCaP"
/lab_host="EMDH10B"
/clone_lib="MAPcL"
/note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkook Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:
Pred. No.: 4.99e-120 Length: 885
Score: 135.00 Matches: 135
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 69.95% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-9 (1-193) x CA453926 (1-885)

Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 86 GAAGGGAAGGACCTCGCTGATCAGAGCTGACTCTGGAGCTGACCCCATGCTGTT 145
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 146 CCTGGAATGTGACCCCTCAGTGTGCTGGCAGCAGCAGTGTCCCTGAGTCTCTCTG 205
Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPheLysIleProCysThr 100
Db 206 AAGGTGGATTAGTTTGGAGAGAGGAGTGGCTGGCTCTCTGATCAAGATCCCATGC 265
Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 266 GACTACATTTGACCTGTACCTTTGAACACTCTGTGATGCTGATGCTGATGCTTAA 325
Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 326 ACTGGGAGCCCTGCCAGAGCCCTGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCT 385
Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
Db 386 AAAGAAGGAACCTACTCACTCCCAAGAGCGAATTCGCTGTGCTGCTGCTGCTGCTG 445
Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSer 175
Db 446 AGTTGGCTCACACCGGGAACCTACCGCATAGAGAGCGTCTCTGAGC 490
RESULT 3
AU280628 559 bp mRNA linear EST 31-JUL-2003
LOCUS AU280628 NCRRM1 Homo sapiens cDNA clone NCRRM1000016 5', mRNA
DEFINITION sequence.
ACCESSION AU280628


```

VERSION AU280628.1 GI:28299855
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 559)
AUTHORS Imabayashi,H., Mori,T., Gojo,S., Kiyono,T., Sugiyama,T., Irie,R.,
         Isogai,T., Hata,J., Tomoya,Y. and Umezawa,A.
TITLE Redifferentiation of dedifferentiated chondrocytes and
        chondrogenesis of human bone marrow stromal cells via chondrosphere
        formation with expression profiling by large-scale cDNA analysis
JOURNAL Exp. Cell Res. 288 (1), 35-50 (2003)
MEDLINE 22760698
PUBMED 12878157
COMMENT Contact: Takao Isogai
        Genomics Laboratory
        Helix Research Institute
        1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
        Tel: 81-438-52-3975
        Fax: 81-438-52-3986
        Email: genomics@hri.co.jp
        HRI human cDNA Project, Sugiyama,T.; Wakamatsu,A.; Irie,R.;
        Umezawa,A.; Fukuma,M.; Kusakari,S.; Hata,J.; Iehii,S.; Yamamoto,J.;
        Isono,Y.; Saito,K.; Nakamura,Y.; Masuho,Y.; Nagai,K.; Isogai,T.
        HRI human cDNA project; cDNA library construction & 5'-end one
        pass sequencing; Helix Research Institute.
FEATURES             Location/Qualifiers
     source           1..559
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NCRRM1000016"
                     /cell_type="embryonal carcinoma"
                     /clone_lib="NCRRM1"
                     /note="Vector: pME18SPL3; mRNA from uninduced embryonal
                     carcinoma"
ORIGIN
Alignment Scores:
Pred. No.:          6.09e-98          Length:          559
Score:              112.00            Matches:         112
Percent Similarity: 100.00%            Conservative:     0
Best Local Similarity: 100.00%          Mismatches:       0
Query Match:        58.03%             Indels:           0
DB:                  1                 Gaps:            0

US-10-030-937-9 (1-193) x AU280628 (1-559)
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 153 GAAGGGAAGGACCCCTGCGGTGATCAGAGCCCTGATCTGGAGCCCTGACCCCATCGTCTGTT 212

QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 213 CCTGGAATGTGACCTCAGTGTCTGTGGGACACACAGTGTCCCTCGAGTTCTCTCTCTG 272

QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
Db 273 AAGGTGGATTAGTTTGTGAGAAGGAGGTGGCTGCTCGATCAAGATCCATCCATGCACA 332

QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 333 GACTACATGGCAGTGTACCTTTGAACACTTCTGTGATGTGCTTGCATGTAAATTCCT 392

QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 393 ACTGGGGAGCCCTGCCAGAGCCCTGCTAGTACTATGGCTTCTTTCCTGCTCCCTTC 452

QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPhe 152
Db 453 AAAGAAGGAACCTACTACTCTGCCCCAAGAGCGAATTC 488

```

```

RESULT 4
BF509172
LOCUS UI-H-B14-aov-C-10-0-UI-s1 NCI CGAP_Sub8 Homo sapiens cDNA clone
DEFINITION IMAGE:3086203 3', mRNA sequence.
ACCESSION BF509172
VERSION BF509172.1 GI:11592470
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 698)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
        Tumor Gene Index
        Unpublished (1997)
        Contact: Robert Strausberg, Ph.D.
        Email: cgapbs-r@mail.nih.gov
        Oligo-dt track not found, Not I site shown in beginning of sequence
        is likely internal to the message. cDNA Library Preparation: M.B.
        Soares Lab Clone distribution: NCI-CGAP clone distribution
        information can be found through the I.M.A.G.E. Consortium/LLNL at:
        www-bio.llnl.gov/bbrp/image/image.html
        Seq primer: M13 Forward
        POLYA=No.
FEATURES             Location/Qualifiers
     source           1..698
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3086203"
                     /lab_host="DH10B (life Technologies)"
                     /clone_lib="NCI CGAP Sub8"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP_Sub8
                     is a subtracted library derived from NCI CGAP_Sub5. The
                     NCI CGAP_Sub8 library had 2.5 million recombinants. A
                     single-stranded DNA preparation of NCI CGAP_Sub5 was used
                     as a tracer in a subtractive hybridization with a driver
                     comprising a pool of clones from NCI CGAP_Sub5 (IMAGE
                     clone ids 2732833-2737415, 3068040-3069191; 25% of the
                     driver population), a pool of clones from NCI CGAP_Sub4
                     (IMAGE clone ids 2723592-2729326; 25% of the driver
                     population), NCI CGAP_Sub6 (pool AIF-AJU, IMAGE ids
                     2728969-2733190; 25% of the driver population), and
                     NCI CGAP_Sub7 (IMAGE ids 3069192-3072238,
                     3081864-3084550; 25% of the driver population).
                     Subtraction was performed as previously described
                     (Bonaldi, Lennon & Soares (1996): Normalization and
                     Subtraction: Two Approaches To Facilitate Gene Discovery.
                     Genome Research 6, 791-806.
                     TAG_SEQ=None found"
ORIGIN
Alignment Scores:
Pred. No.:          7.52e-98          Length:          698
Score:              112.00            Matches:         152
Percent Similarity: 98.70%            Conservative:     0
Best Local Similarity: 98.70%          Mismatches:       1
Query Match:        58.03%             Indels:           2
DB:                  2                 Gaps:            0

US-10-030-937-9 (1-193) x BF509172 (1-698)
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 170 GAAGGGAAGGACCCCTGCGGTGATCAGAGCCCTGATCTGGAGCCCTGACCCCATCGTCTGTT 229

QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 230 CCTGGAAATGTGACCTCAGTGTCTGTGGGACACACAGTGTCCCTCGAGTTCTCTCTCTG 289

```

QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPileLysLeuProCysThr 100
 |||||
 Db 290 AAGGTGAATTAGTTTGGAGAGAGGCTGCTGCTCTGGATCAAGATCCCATGCACA 349
 |||||
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 |||||
 Db 350 GACTACATTTGGACCTGACTTTGAACACTTCTGTGATGCTTGACATGTTAATTCCT 409
 |||||
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 |||||
 Db 410 ACTGGGAGCCCTGCCCCAGAGCCCTGGTACCTATGCGCTTCTTCCACTGTCCCTTC 469
 |||||
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPro 160
 |||||
 Db 470 AAAGAGGAACCTACTCACTGCCCAAGAGGAAATTCGT-TGTGCTGACCTGGAGCTGCC 528
 |||||
 QY 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
 |||||
 Db 529 CAGTTGGCTCACACACCGGAACTACCCATAGAGAGGCTCTGAGCAGCAGTGGGAAGCG 588
 |||||
 QY 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 |||||
 Db 589 TCTGGGCTGCATCAAGATCGTGCCTCTCTAAAGGGCATA 628
 |||||

RESULT 5
 BG479322 711 bp mRNA linear EST 21-MAR-2001
 LOCUS 602526351F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4649869 5',
 DEFINITION mRNA sequence.
 ACCESSION BG479322 GI:13411601
 VERSION BG479322.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 711)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHCMI430 row: f column: 14
 High quality sequence stop: 711.
 Location/Qualifiers
 1..711
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4649869"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pOTB1; Site 1: XhoI;
 Site 2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

FEATURES

source

ORIGIN
 Alignment Scores:
 Pred. No.: 7,666-98 Length: 711
 Score: 112.00 Matches: 152

Percent Similarity: 98.70% Conservative: 0
 Best Local Similarity: 98.70% Mismatches: 1
 Query Match: 58.03% Indels: 2
 DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BG479322 (1-711)

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 |||||
 Db 70 GAAGGGAGGACCCCTGCGGTGATCAGAAGGCTGACTCTGGAGCCTGACCCCATCGTGT 129
 |||||
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 |||||
 Db 130 CCTGGAATGTGACCCCTCAGTGTCTGGGAGACACAGTGTCCCTTGATTTCTCTCTG 189
 |||||
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPileLysLeuProCysThr 100
 |||||
 Db 190 AAGGTGAATTAGTTTGGAGAGGAGTGGTGGCTCTGGATCAAGATCCCATGCACA 249
 |||||
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 |||||
 Db 250 GACTACATTTGGACCTGACTTTGAACACTTCTGTGATGCTTGACATGTTAATTCCT 309
 |||||
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 |||||
 Db 310 ACTGGGAGCCCTGCCCCAGAGCCCTGGTACCTATGCGCTTCTTCCACTGTCCCTTC 369
 |||||
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPro 160
 |||||
 Db 370 AAAGAGGAACCTACTCACTGCCCAAGAGGAAATTCGT-TGTGCTGACCTGGAGCTGCC 428
 |||||
 QY 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
 |||||
 Db 429 CAGTTGGCTCACACACCGGAACTACCCATAGAGAGGCTCTGAGCAGCAGTGGGAAGCG 488
 |||||
 QY 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 |||||
 Db 489 TCTGGGCTGCATCAAGATCGTGCCTCTCTAAAGGGCATA 528
 |||||

RESULT 6

BI856212

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LHCMI1998 row: h column: 10

High quality sequence start: 2

High quality sequence stop: 733.

Location/Qualifiers

1..733

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5391801"

/tissue_type="mammary adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_87"

/note="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 7,88e-98 Length: 733
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservatives: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BI856212 (1-733)

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 140 GAAAGGAAAGACCTCGCGTGTATCAGAGCCTGACTCTGGAGCCTGACCCCATCGCGTT 199
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerProLeu 80
Db 200 CTGGAATGTGACCTCAGTGTGGGACACAGTGCCCTGAGTTCCTCTCTG 259
QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
Db 260 AAGGTGGATTTAGTTTGGAGAAGGAGGTGGCTGGCTCTGGATCAGATCCATGCACA 319
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 320 GACTACATTGGCAGCTGACCTTTGAACACTTCTGTGATGTGCTTGACATGTAATTCCT 379
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 380 ACTGGGAGCCTGCGCCAGAGCCCTCGCTACTATGGCTTCCTTGGCCACTGTCCCTTC 439
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
Db 440 AAAGAGGAGGACCTACTACTGCGCCAGAGCGAATTCGT-TGTGCCTGACTGGAGCTGCC 498
QY 160 sSerTrpLeuThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
Db 499 CAGTTGGCTCACACCGGGAACCTACCGCATAGAGAGCGTCTGAGCAGCAGTGGGAAGCG 558
QY 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 559 TCTGGGCTGCATCAAGATCGCTCTCTCTAAAGGGCATA 598

RESULT 7
BI838554 784 bp mRNA linear EST 04-OCT-2001
LOCUS 603086219F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5225092 5',
DEFINITION mRNA sequence.

ACCESSION BI838554

VERSION 1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 784)

NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAM1566 row: f column: 05

High quality sequence stop: 772.

FEATURES

Location/Qualifiers
source 1..784

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5225092"

/lab_host="DH10B"

/clone_lib="NIH_MGC_120"

/note="Organ: pooled pancreas and spleen; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 8,41e-98 Length: 784
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservatives: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BI838554 (1-784)

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 125 GAAAGGAAAGACCTCGCGTGTATCAGAGCCTGACTCTGGAGCCTGACCCCATCGCGTT 184
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 185 CTGGAATGTGACCTCAGTGTGGGACACAGTGCCCTGAGTTCCTCTCTG 244
QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
Db 245 AAGGTGGATTTAGTTTGGAGAAGGAGGTGGCTGGCTCTGGATCAGATCCATGCACA 304
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 305 GACTACATTGGCAGCTGACCTTTGAACACTTCTGTGATGTGCTTGACATGTAATTCCT 364
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 365 ACTGGGAGCCTGCGCCAGAGCCCTCGCTACTATGGCTTCCTTGGCCACTGTCCCTTC 424
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
Db 425 AAAGAGGAGGACCTACTACTGCGCCAGAGCGAATTCGT-TGTGCCTGACTGGAGCTGCC 483
QY 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
Db 484 CAGTTGGCTCACACCGGGAACCTACCGCATAGAGAGCGTCTGAGCAGCAGTGGGAAGCG 543
QY 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 544 TCTGGGCTGCATCAAGATCGCTCTCTCTAAAGGGCATA 583

RESULT 8

BI091220

LOCUS

DEFINITION

602856051F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4997365 5',

mRNA sequence.

ACCESSION BI091220

BI091220 842 bp mRNA linear EST 20-JUN-2001

602856051F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4997365 5',

mRNA sequence.

ACCESSION BI091220

```

VERSION      BI091220.1  GI:14509550
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 842)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1024 row: e column: 14
High quality sequence stop: 739.
Location/Qualifiers
1. .842
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4997365"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_lib="NIH_MGC 10"
/note="Organ: Cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

ORIGIN
source
Alignment Scores:
Pred. No.: 9e-98 Length: 842
Score: 112.00 Matches: 147
Percent Similarity: 98.66% Conservativeness: 0
Best Local Similarity: 98.66% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BI091220 (1-842)

Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 132 GAAGGGAAGACCTCGCGTGATCAGAGCCTGACTCTGGAGCCTGACCCATCGTCTG 191
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 192 CCTGGAATGTGACCTCAGTGTGCTGGGAGCAGCAGTGTCCCTGAGTCTCTCTCTG 251
Qy 81 LysValAspLeuValLeuGluValAlaGlyLeuThrIleValIleProCysThr 100
Db 252 AAGTGGATTAGTATTTGGAGAGAGGTGCTGGCTCTGGATCAAGATCCCATGCACA 311
Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 312 GACTACATTGSCAGCTGTACCTTTGAACACTTCTGTGATGTGCTTGACATGTATTCCT 371
Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 372 ACTGGGAGACCTCGCCAGAGCCCTCGGTACCTATGGGCTTCCTGCCACTGTCCTTC 431
Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
Db 432 AAAGAGGAACCTACTACTACCCCAAGAGCGAATTCGT-TGTGCTGTACCTGGAGCTGCC 490
Qy 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
Db 491 CAGTTGGCTACCCACCGGAACTACCGCATAGAGAGGCTCTCTGAGCAGCAGTGGGAGCG 550

```

```

Qy 180 gLeuGlyCysIleLysIleAlaAla 188
Db 551 TCTGGGCTGCATCAAGATCGCTGCT 575

RESULT 9
BI020051
LOCUS     603037236F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5178273 5',
DEFINITION mRNA sequence.
ACCESSION BI020051
VERSION   BI020051.1 GI:15931601
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 858)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1144 row: g column: 10
High quality sequence stop: 706.
Location/Qualifiers
1. .858
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5178273"
/lab_host="DH10B"
/clone_lib="NIH_MGC 115"
/note="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 9.16e-98 Length: 858
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservativeness: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BI020051 (1-858)

Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 128 GAAGGGAAGACCTCGCGTGATCAGAGCCTGACTCTGGAGCCTGACCCATCGTCTG 187
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 188 CCTGGAATGTGACCTCAGTGTGCTGGGAGCAGCAGTGTCCCTGAGTCTCTCTCTG 247
Qy 81 LysValAspLeuValLeuGluValAlaGlyLeuThrIleValIleProCysThr 100
Db 248 AAGTGGATTAGTATTTGGAGAGAGGTGCTGGCTCTGGATCAAGATCCCATGCACA 307

```

```

QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 308 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGCTTGACATGTTAATTCT 367
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 368 ACTGGGAGCCCTGCCAGAGCCCTGGTACCTATGGCTTCTTGCCACTGTGCCCTTC 427
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
Db 428 AAAGAAGGAACCTACTACTGCCCCAAGAGCGAATTGCT-TGTGCTGTGACCTGGAGCTGCC 486
QY 160 oSerTripleLeuThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
Db 487 CAGTTGGCTCACCCAGCGGAACCTACCGCATAGAGAGCGTCTTGAGCAGCAGTGGGAAGCG 546
QY 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 547 TCTGGGCTGCATCAAGATCGCTGCTCTCTTAAGGGGCATA 586

RESULT 10
BQ676659
LOCUS BQ676659 895 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT 8211756 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6258669
5', mRNA sequence.
ACCESSION BQ676659
VERSION BQ676659.1 GI:21789338
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 895)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DPF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2414 row: 0 column: 22
High quality sequence stop: 567.
FEATURES
Location/Qualifiers
1..895
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6258669"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/notice="Organ: skin; Vector: pOTE7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into ECORI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 9.54e-98 Length: 895
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservative: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2

```

```

DB: 5 Gaps: 0
US-10-030-937-9 (1-193) x BQ676659 (1-895)
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 32 GAAGGAGAGCCCTGCGGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTCGT 91
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 92 CTGGAATGTGACCTCAGTGTCTGTGGCAGCAGCAGTCTCCCTGAGTCTCTCTCTG 151
QY 81 LysValAspLeuValLeuLysGluValIleGlyLeuThrIleLysIleProCysThr 100
Db 152 AAGGTGGATTAGTGTGGAGAGGAGGTGCTGCGCTCTGGATCAAGATCCCATGCACA 211
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 212 GACTACATTGGAGCTGTACCTTTGAACACTTCTGTGATGCTTGACATGTTAATTCT 271
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 272 ACTGGGAGCCCTGCCAGAGCCCTGCTGCTATGGCTTCTTGCACCTGTCCCTTC 331
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
Db 332 AAAGAAGGAACCTACTACTGCCCAAGAGCGAATTGCT-TGTGCTGTGACCTGGAGCTGCC 390
QY 160 oSerTripleLeuThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
Db 391 CAGTTGGCTCACCCAGCGGAACCTACCGCATAGAGAGCGTCTTGAGCAGCAGTGGGAAGCG 450
QY 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 451 TCTGGGCTGCATCAAGATCGCTGCTCTCTTAAGGGGCATA 490

RESULT 11
AL560604 910 bp mRNA linear EST 02-APR-2004
LOCUS AL560604 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CSODL003YG14 5-PRIME, mRNA sequence.
ACCESSION AL560604
VERSION AL560604.3 GI:46185974
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 910)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:31284734.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
3172.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CSODL003BD07QPI&c=3172.f.
FEATURES
Location/Qualifiers
1..910
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODL003YG14"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"

```

/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
Pred. No.: 9.69e-98 Length: 910
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservative: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 1 Gaps: 0

US-10-030-937-9 (1-193) x AL560604 (1-910)

Qy 41 GluGlyLysAspProAlaValIleAtrqSerLeuThrLeuGluProAaspProIleValVal 60
Db 143 GAAGGGAAGGACCTCGCGGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTCT 202
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerProLeu 80
Db 203 CCTGGAATGTGACCTCAGTGTGCGGCAGCAGCAGTGTCCCTGAGTTCCTCTG 262
Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPheLysIleProCysThr 100
Db 263 AAGGTGATTTAGTTTGGAGAGAGGAGTGGTGGCTCTGGATCAAGATCCCATGCACA 322
Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 323 GACTACATTTGACCTGACTTTGAACACTTCTGTGATGCTTGACATGTTAATTCCT 382
Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 383 ACTGGGAGCCCTGCGCCAGAGCCCTCGGTACCTATGCGTCTCTGCCACTGTCCCTTC 442
Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
Db 443 AAAGAAGAAGCACTACTCAGTCCCAAGAGCAATTCGT-TGTGCTTACCTGGAGCTGCC 501
Qy 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
Db 502 CAGTTGGCTACACCGGGAACCTACCCATAGAGAGCGTCTCTGAGAGCAGTGGGAAGCG 561
Qy 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 562 TCTGGGCTGCATCAAGATCGTGCCTCTCTAAAGGGCATA 601

RESULT 12
BUI51364
LOCUS
DEFINITION AGENCOURT 8119275 Lupski dorsal root ganglion Homo sapiens EST 03-SEP-2002
ACCESSION BUI51364
VERSION BUI51364.1 GI:22664896
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 937)
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: ccapbs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM13561 row: b column: 07
High quality sequence stop: 684.

FEATURES

source 1..937
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6179622"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/clone_lib="Lupski_dorsal_root_ganglion"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site 1:
NotI; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGTCCG-3' and
5'-GACTAGTTCTAGATCGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."

ORIGIN

Alignment Scores:
Pred. No.: 9.97e-98 Length: 937
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservative: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 5 Gaps: 0

US-10-030-937-9 (1-193) x BUI51364 (1-937)

Qy 41 GluGlyLysAspProAlaValIleAtrqSerLeuThrLeuGluProAaspProIleValVal 60
Db 194 GAAGGGAAGGACCTCGCGGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTCT 253
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 254 CCTGGAATGTGACCTCAGTGTGCGGCAGCAGCAGTGTCCCTGAGTTCCTCTG 313
Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPheLysIleProCysThr 100
Db 314 AAGGTGATTTAGTTTGGAGAGAGGAGTGGTGGCTCTGGATCAAGATCCCATGCACA 373
Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 374 GACTACATTTGACCTGACTTTGAACACTTCTGTGATGCTTGACATGTTAATTCCT 433
Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 434 ACTGGGAGCCCTGCGCCAGAGCCCTCGGTACCTATGCGTCTCTGCCACTGTCCCTTC 493
Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160
Db 494 AAAGAAGAAGCACTACTCAGTCCCAAGAGCAATTCGT-TGTGCTTACCTGGAGCTGCC 552
Qy 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180
Db 553 CAGTTGGCTACACCGGGAACCTACCCATAGAGAGCGTCTCTGAGAGCAGTGGGAAGCG 612
Qy 180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 613 TCTGGGCTGCATCAAGATCGTGCCTCTCTAAAGGGCATA 652

RESULT 13

BUI56606
LOCUS
DEFINITION AGENCOURT_10188553 NIH_MGC_109 Homo sapiens cdna clone
948 bp mRNA linear EST 16-SEP-2002

```

IMAGE:6584324 5', mRNA sequence.
ACCESSION BUS56606 Length: 948
VERSION BUS56606.1 GI:22906878 /mol_type="mRNA"
KEYWORDS EST. /db_xref="taxon:9606"
SOURCE Homo sapiens /clone="IMAGE:6584324"
ORGANISM Homo sapiens (human) /lab_host="DH10B (phage-resistant)"
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; /note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. XhoI; cDNA made by oligo-dr priming. Directionally cloned
NIH-MGC http://mgc.nci.nih.gov/. into EcoRI/XhoI sites using the following 5' adaptor:
National Institutes of Health, Mammalian Gene Collection (MGC) GGACGAG(G). Library constructed by Ling Hong in the
Unpublished (1999) laboratory of Gerald M. Rubin (University of California,
Contact: Robert Strausberg, Ph.D. Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Email: cgapbs-remail.nih.gov Superscript II RT (Life Technologies). Note: this is a
Tissue Procurement: ATCC NIH_MGC Library."
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: LUCM2794 row: h column: 20
High quality sequence stop: 645.
FEATURES
Location/Qualifiers
1. 948
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6584324"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"
/note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dr priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 1.01e-97 Length: 948
Score: 112.00 Matches: 149
Percent Similarity: 98.68% Conservative: 0
Best Local Similarity: 98.68% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 5 Gaps: 0

US-10-030-937-9 (1-193) x BUS56606 (1-948)

Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 143 GAAGGGAAGGACCCCTGGGTGATCAGAGACCTGACTCTGGAGCCTGACCCCATGCTGCTT 202
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 203 CCTGGAAATGTGACCCCTCAGTGTGCTGGGACGACCCAGTGTCCCTGAGTTCCTCTG 262
Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrIleLysIleProCysThr 100
Db 263 AAGGTGGATTTAGTTTGGAGAAGAGAGTGCTGGCCCTCTGGATCAAGATCCCATGCACA 322
Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 323 GACTACATTTGGCAGCTGTACCTTTGAACACTTCTGTGATGCTGACATGTTAATTCCT 382
Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 383 ACTGGGAGCCCTGCCAGAGCCCTGGCTACCTATGGCTTCCTTGGCACTGTCCCTTC 442
Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla-ValProAspLeuGluLeuPr 160

IMAGE:6584324 5', mRNA sequence.
ACCESSION BUS56606 Length: 949 bp mRNA linear EST 29-MAR-2002
VERSION BUS56606.1 GI:22906878 /mol_type="mRNA"
KEYWORDS EST. /db_xref="taxon:9606"
SOURCE Homo sapiens /clone="IMAGE:6584324"
ORGANISM Homo sapiens (human) /lab_host="DH10B (phage-resistant)"
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; /note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. XhoI; cDNA made by oligo-dr priming. Directionally cloned
NIH-MGC http://mgc.nci.nih.gov/. into EcoRI/XhoI sites using the following 5' adaptor:
National Institutes of Health, Mammalian Gene Collection (MGC) GGACGAG(G). Size-selected >500bp for average insert size
Unpublished (1999) of Gerald M. Rubin (University of California, Berkeley)
Contact: Robert Strausberg, Ph.D. using ZAP-cDNA synthesis kit (Stratagene) and
Email: cgapbs-remail.nih.gov Superscript II RT (Life Technologies). Note: this is a NIH_MGC
Tissue Procurement: Lou Staudt Library."
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: LUCM2072 row: 1 column: 20
High quality sequence stop: 741.
FEATURES
Location/Qualifiers
1. 949
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6584324"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_99"
/note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
XhoI; cDNA made by oligo-dr priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGACGAG(G). Size-selected >500bp for average insert size
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN
Alignment Scores:
Pred. No.: 1.01e-97 Length: 949
Score: 112.00 Matches: 152
Percent Similarity: 98.70% Conservative: 0
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 58.03% Indels: 2
DB: 5 Gaps: 0

US-10-030-937-9 (1-193) x BQ060062 (1-949)

Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 111 GAAGGGAAGGACCCCTGGGTGATCAGAGACCTGACTCTGGAGCCTGACCCCATGCTGCTT 170
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80

```

Db 171 CCTGGAATGTGACCTCAGTGTGCTGGGAGCAGCAGTGTCCCTCAGTTCCTCTCTG 230
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysLeuProCysThr 100
 Db 231 AAGGTGGATTAGTTTGGAGAGGAGGTGGCTCTGGATCAAGATCCCATGCACA 290
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 291 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGTGATCAAGATCCCATGCACA 350
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 351 ACTGGGAGCCCTGGCCAGAGCCCTGCGTACCTAATGGGCTTCTTGCCACTGTCCCTTC 410
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla ValProAspLeuGluLeuPro 160
 Db 411 AAAGAAGGAACCTACTCACTGCCCAAGAGCGAATTCGT-TGTGCTGACCTGGAGCTGCC 459
 QY 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysAr 180
 Db 470 CAGTTGGCTCACCACCGGAACTACCCATAGAGAGCGTCTGAGCAGCAGTGGGAAGCG 529
 QY 180 gLeuGlyCysIleLysIleAlaLaserLeuLysGlyIle 193
 Db 530 TCTGGGCTGCATCAAGATCGCTGCTCTTAAAGGGCATA 569

RESULT 15
 BM474816
 LOCUS
 DEFINITION AGENCOURT 6476614 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562559
 S', mRNA sequence.
 ACCESSION BM474816
 VERSION BM474816.1 GI:18523858
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 973)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-femail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12292 row: c column: 08
 High quality sequence stop: 603.

FEATURES
 source 1..973
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5562559"
 /tissue types="duodenal adenocarcinoma, cell line"
 /lab host="DH10B (phage-resistant)"
 /clone lib="NIH_MGC_88"
 /note="Organ: small intestine; Vector: pCMV-SPORT6;
 Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
 oligo-dr primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores: 1.03e-97 Length: 973
 Pred. No.: 112.00 Matches: 149
 Score:

Percent Similarity: 98.68% Conservative: 0
 Best Local Similarity: 98.68% Mismatches: 1
 Query Match: 58.03% Indels: 2
 DB: 4 Gaps: 0
 US-10-030-937-9 (1-193) x BM474816 (1-973)
 QY 41 GluGlyValAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 118 GAAGGGAAGGACCTCGGGTGATCAAGAGCCTGACTCTGGAGCCTGACCCCATCGTCT 177
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 Db 178 CCTGGAAATGTGACCTCAGTGTCTGGGGAGCAGCAGTGTCCCTGAGTTCCTCTG 237
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
 Db 238 AAGGTGGATTAGTTTGGAGAGGAGGTGGCTCTGGATCAAGATCCCATGCACA 297
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 298 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGTGATCAAGATCCCATGCACA 357
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 358 ACTGGGAGCCCTGCCAGAGCCCTGCGTACCTATGGGCTTCTTGCCACTGTCCCTTC 417
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAla ValProAspLeuGluLeuPro 160
 Db 418 AAAGAAGGAACCTACTCACTGCCCAAGAGCGAATTCGT-TGTGCTGACCTGGAGCTGCC 476
 QY 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysAr 180
 Db 477 CAGTTGGCTCACCACCGGAACTACCCATAGAGAGCGTCTGAGCAGCAGTGGGAAGCG 536
 QY 180 gLeuGlyCysIleLysIleAlaLaserLeu 190
 Db 537 TCTGGGCTGCATCAAGATCGCTGCCCTCTCTA 567

Search completed: November 19, 2004, 05:51:07
 Job time : 3325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 17:56:16 ; Search time 7857.24 Seconds
(without alignments)
1161.592 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 1018

Sequence: 1 MQSMDPLIALGILLATP.....LSSGKRLGCIKIAASLKGI 193

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2 1/USPTO.spool/p/US10030937/runat.16112004.153013.2913/app.query.fasta_1.789
-DB=GenEmbl -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOPCU=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -FRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=spt -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10030937.cgn 1 1 7771 @runat.16112004.153013.2913 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1005	98.7	1043	9 HUMGM2A	L01439 Human GM2-a
2	1005	98.7	2436	6 AX330938	AX330938 Sequence
3	1005	98.7	2436	9 HSGM2AP	X62078 H.sapiens m
4	1000	98.2	953	9 HUMGM2	M76477 Human G-M2

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1	HUMGM2A	Human GM2-activator protein (GM2A)	1043 bp	mRNA	linear	PRI 09-NOV-1994
LOCUS	HUMGM2A	Human GM2-activator protein (GM2A)	1043 bp	mRNA	complete cds.	
DEFINITION	L01439	Human GM2-activator protein (GM2A)	1043 bp	mRNA	complete cds.	
ACCESSION	L01439	Human GM2-activator protein (GM2A)	1043 bp	mRNA	complete cds.	
VERSION	L01439.1	GI:183358				
KEYWORDS	GM2 activator protein.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
AUTHORS	Xie, B., Kennedy, J.L., McInnes, B., Auger, D. and Mahuran, D.					
TITLE	Identification of a processed pseudogene related to the functional gene encoding the GM2 activator protein: localization of the pseudogene to human chromosome 3 and the functional gene to human chromosome 5					
JOURNAL	Genomics 14 (3), 796-798 (1992)					
MEDLINE	93052421					
PUBMED	1427911					
COMMENT	Original source text: Homo sapiens cDNA to mRNA.					
FEATURES	location/Qualifiers					
source	1. .1043					
organism	/organism="Homo sapiens"					
/mol type	/mol type="mRNA"					
/db xref	/db xref="taxon:9606"					
/map	/map="5"					
gene	1. .1043					

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS
Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE
Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL
Patent: WO 0194629-A 1447 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Location/Qualifiers
1..2436
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

```

ORIGIN
Alignment Scores:
Pred. No.: 7 19e-87 Length: 2436
Score: 1005.00 Matches: 191
Percent Similarity: 98.96% Conservative: 0
Best Local Similarity: 98.96% Mismatches: 2
Query Match: 98.72% Indels: 0
DB: 6 Gaps: 0
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

```

DB: 6 Gaps: 0

US-10-030-937-9 (1-193) x AX330938 (1-2436)

Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20

Db 59 ATGCAGTCCCTGATGCAGGCTCCCTCTCTGATCGCCCTGGGCTTGCTTCGCGAGCCCT 118

Qy 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40

Db 119 GCCAAGCCCACTGAAAAGCCATCCAGCTCAGTAGCTTTTCCTGGGATAACTGTGAT 178

Qy 41 GluGlyLysAspProLalaValIleArgSerLeuThrLeuGluProAspProIleValVal 60

Db 179 GAAGGGAAGACCCCTCGCGTGATCAGAAGCCCTGACTCTCGAGGCTGACCCCATCGTCGTT 238

Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80

Db 239 CCTGGAAATGTGACCTCAGTGTCGGGCGACGACAGGTGCCCTCGATGTTCTCTCTG 238

Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100

Db 299 AAGGTGGATTAGTTTGGAGAGGAGGTGGCTCTGGATCAAGATCCCATGCACA 358

Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120

Db 359 GACTACATTGGCAGCTGTACCTTTGACACTTCTGTGATGTGCTTGACATGTTAATTCCT 418

Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProphe 140

Db 419 ACTGGGAGGCCCTGCCAGAGCCCTCGCTACCTATGGGCTTCCTTTGCCACTGTCCTTC 478

Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160

Db 479 AAAGAGGACCTACTCACTGCCCAAGAGCGAATTCGTTGTGCTGACCTGGAGCTGCC 538

Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180

Db 539 AGTTGGCTCACCACCGGAACTACCGCATAGAGCGCTCTTGAGCAGCAGTGTGGAGCGT 598

Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193

Db 599 CTGGGCTGCATCAAGATCGCTGCCTCTCTAAAGGGCATA 637

Db	599	CTGGCTGCATCAGATCGCTGCTCTCTCTAAAGGGCATA	637
RESULT 3			
HSGM2APT		HSGM2APT	2436 bp
LOCUS		H.sapiens mRNA for GM2 activator protein.	linear
DEFINITION			
ACCESSION		X62078	
VERSION		X62078.1	GI:313158
KEYWORDS		G(M2) activator protein.	
			PRI 15-FEB-1995

```

SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE    1 (bases 1 to 2436)
AUTHORS     Klimah,H., Tanaka,A., Schnabel,D., Nakano,T., Schroder,M., Suzuki,K.
            and Sandhoff,K.
TITLE       Characterization of full-length cDNAs and the gene coding for the
            human GM2 activator protein
JOURNAL     FEBS Lett. 289 (2), 260-264 (1991)
MEDLINE     92008637
PUBMED      1915857
REFERENCE    2 (bases 1 to 2436)
AUTHORS     Klimah,H., Klein,A., van Echten,G., Schwarzmamm,G., Suzuki,K. and
            Sandhoff,K.
TITLE       Over-expression of a functionally active human GM2-activator
            protein in Escherichia coli
JOURNAL     Biochem. J. 292 (Pt 2), 571-576 (1993)
MEDLINE     93277527
PUBMED      8503891
FEATURES    Location/Qualifiers
             1..2436
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /isolate="patient with juvenile form of Sandhoff disease"
             /db_xref="taxon:9606"
             /clone="pUC18"
             /cell_type="fibroblast"
             /clone_lib="cDNA"
             59..640
             /note="alternative"
             /codon_start=1
             /product="GM2 activator protein"
             /protein_id="CAA43993.1"
             /db_xref="GI:673415"
             /db_xref="GOA:P17900"
             /translation="MQSLMQAPLLIALGLLIATPAQHLKPKSLSFSWNCDEGKDP
             PAVIRSLTLEPDPPIVPGNVTLSVVGSTSVPLSKVDLVLEKEVAGLWIKIPCTDY
             IGSCTFEHFCVDLMDLIPTEGPCPEPLRTYGLPCHPCPKFGTYSLPKSEFVVPDLELP
             SWLTGNRYIESVLSSSGKRLGCIKIAASLKI"
             71..640
             /note="alternative"
             /codon_start=1
             /product="GM2 activator protein"
             /protein_id="CAA43994.1"
             /db_xref="GI:673416"
             /translation="MQAPLLIALGLLIATPAQHLKPKSLSFSWNCDEGKDP
             RSLTLEPDPPIVPGNVTLSVVGSTSVPLSKVDLVLEKEVAGLWIKIPCTDY
             TFEHFCVDLMDLIPTEGPCPEPLRTYGLPCHPCPKFGTYSLPKSEFVVPDLELP
             TGNRYIESVLSSSGKRLGCIKIAASLKI"
CDS
FEATURES    source
             1..2436
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /isolate="patient with juvenile form of Sandhoff disease"
             /db_xref="taxon:9606"
             /clone="pUC18"
             /cell_type="fibroblast"
             /clone_lib="cDNA"
             59..640
             /note="alternative"
             /codon_start=1
             /product="GM2 activator protein"
             /protein_id="CAA43993.1"
             /db_xref="GI:673415"
             /db_xref="GOA:P17900"
             /translation="MQSLMQAPLLIALGLLIATPAQHLKPKSLSFSWNCDEGK
             PAVIRSLTLEPDPPIVPGNVTLSVVGSTSVPLSKVDLVLEKEVAGLWIKIPCTDY
             IGSCTFEHFCVDLMDLIPTEGPCPEPLRTYGLPCHPCPKFGTYSLPKSEFVVPDLELP
             SWLTGNRYIESVLSSSGKRLGCIKIAASLKI"
             71..640
             /note="alternative"
             /codon_start=1
             /product="GM2 activator protein"
             /protein_id="CAA43994.1"
             /db_xref="GI:673416"
             /translation="MQAPLLIALGLLIATPAQHLKPKSLSFSWNCDEGKDP
             RSLTLEPDPPIVPGNVTLSVVGSTSVPLSKVDLVLEKEVAGLWIKIPCTDY
             TFEHFCVDLMDLIPTEGPCPEPLRTYGLPCHPCPKFGTYSLPKSEFVVPDLELP
             TGNRYIESVLSSSGKRLGCIKIAASLKI"
CDS
FEATURES    source
             1..953
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             91..672
             /codon_start=1
             /product="G-M2 activator protein"
             /protein_id="AAA35907.1"
             /db_xref="GI:183357"
             /translation="MQSLMQAPLLIALGLLIATPAQHLKPKSLSFSWNCDEGK
             PAVIRSLTLEPDPPIVPGNVTLSVVGSTSVPLSKVDLVLEKEVAGLWIKIPCTDY
             IGSCTFEHFCVDLMDLIPTEGPCPEPLRTYGLPCHPCPKFGTYSLPKSEFVVPDLELP
             SWLTGNRYIESVLSSSGKRLGCIKIAASLKI"
             91..159
             sig_peptide
             160..669
             mat_peptide
             /product="G-M2 activator protein"
ORIGIN
Alignment Scores:
Pred. No.: 7.49e-87 Length: 953
Score: 1000.00 Matches: 190
Percent Similarity: 98.45% Conservative: 0
Query Match: 98.72% Mismatches: 3
DB: 9
US-10-030-937-9 (1-193) x HSGM2APT (1-2436)
Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db 59 ATGCAGTCCCTGATGACGGCTCCCTCTGATCGCCCTGGGCTTCTTCGCGACCCCT 118
Qy 21 AlaGlnAlaHisLeuLeuLysLeuProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
Db 119 GCGCAAGCCCACTGAAAGAACCCATCCAGCTCAGTAGCTTTCTCGGATAACTGTGAT 178
Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60

```


Best Local Similarity: 98.45% Mismatches: 3
 Query Match: 98.23% Indels: 0
 DB: 9 Gaps: 0

US-10-030-937-9 (1-193) x BC009273 (1-2413)

Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 Db 30 ATGCAGTCCCTGATGCGAGGCTCCCTCTGATCGCCCTGGGCTTGTCTCGCGGCCCT 89
 Qy 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 Db 90 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTCTCTGGGATACTGTGAT 149
 Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 150 GAAGGGAAGGACCTCGCGGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTGT 209
 Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 Db 210 CTGGAAATGTGACCTCAGTGTCTGGGAGCAGCAGTGTCTCCCTGAGTCTCTCTG 269
 Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
 Db 270 AAGTGGATTAGTTTGGAGAGGAGGTGGCTGCTCTGTGATCAAGATCCCATGCACA 329
 Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 330 GACTACATTGGCAGCTGTACCTTTGAACACTCTGTGATGTGCTGACATGTTAATTCCT 389
 Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 390 ACTGGGAGCCTCCAGAGCCTCGTACCTATGAGCTTCTGTGCTGCTGCTGCTGCTGCT 449
 Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 Db 450 AAAGAAGAACCTACTACTGCTCCCAAGAGCGAATTCGTGTGCTGCTGCTGCTGCTGCT 509
 Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 Db 510 AGTTGGCTCACCACCGGAACTACCGCATAGAGCGCTCTGAGCAGCAGTGGGAAGCGT 569
 Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db 570 CTGGGCTGCATCAAGATCGCTGCTCTCTAAAGGGCATA 608

RESULT 6
 CQ728078 1045 bp DNA linear PAT 03-FEB-2004
 LOCUS
 DEFINITION Sequence 14012 from Patent WO02068579.
 CQ728078
 ACCESSION
 VERSION CQ728078.1 GI:42295943
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
 Kits, such as nucleic acid arrays, comprising a majority of
 human exons or transcripts, for detecting expression and other uses
 thereof
 Parent: WO 02068579-A 14012 06-SEP-2002;
 PE Corporation (NY) (US)
 JOURNAL
 FEATURES
 Location/Qualifiers
 1..1045
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Alignment Scores: 2.02e-86 Length: 1045
 Pred. No.:

Score: 996.00 Matches: 188
 Percent Similarity: 98.45% Conservative: 2
 Best Local Similarity: 97.41% Mismatches: 3
 Query Match: 97.84% Indels: 0
 DB: 6 Gaps: 0

US-10-030-937-9 (1-193) x CQ728078 (1-1045)

Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 Db 92 ATGCAGTCCCTGATGCGAGGCTCCCTCTGATCGCCCTGGGCTTGTCTCGCGGCCCT 151
 Qy 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 Db 152 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTCTCTGGGATACTGTGAT 211
 Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 212 GAAGGGAAGGACCTCGCGGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTGT 271
 Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 Db 272 CTGGAAATGTGACCTCAGTGTCTGGGAGCAGCAGTGTCTCCCTGAGTCTCTCTG 331
 Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
 Db 332 AAGTGGATTAGTTTGGAGAGGAGGTGGCTGCTCTGTGATCAAGATCCCATGCACA 391
 Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 392 GACTACATTGGCAGCTGTACCTTTGAACACTCTGTGATGTGCTGATGCTGATGCTGAT 451
 Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 452 ACTGGGAGCCTCCAGAGCCTCGTACCTATGAGCTTCTGTGCTGCTGCTGCTGCTGCT 511
 Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 Db 512 AAAGAAGAACCTACTACTGCTCCCAAGAGCGAATTCGTGTGCTGCTGCTGCTGCTGCT 571
 Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 Db 572 AGTTGGCTCACCACCGGAACTACCGCATAGAGCGCTCTGAGCAGCAGTGGGAAGCGT 631
 Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db 632 CTGGGCTGCATCAAGATCGCTGCTCTCTAAAGGGCATA 670

RESULT 7
 HSGM2APB 1093 bp mRNA linear PRI 10-APR-1992
 LOCUS
 DEFINITION H.sapiens RNA for GM2-activator protein (clones pGAP2 & pGAP3).
 CQ728078
 ACCESSION
 VERSION X61095
 KEYWORDS
 SOURCE G(M2) activator protein.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 (bases 1 to 1093)
 Nagajaran, S., Chen, H.C., Li, S.C., Li, Y.T. and Lockyer, J.M.
 Evidence for two cDNA clones encoding human GM2-activator protein
 Biochem. J. 282 (Pt 3), 807-813 (1992)
 JOURNAL
 MEDLINE
 PUBMED
 154364
 2 (bases 1 to 1093)
 Lockyer, J.
 Direct Submission
 Submitted (26-JUL-1991) J. Lockyer, Tulane University Medical
 school, Human Genetics Program, 1430 Tulane Ave., New Orleans LA
 70112, USA
 FEATURES
 Location/Qualifiers
 1..1093
 source

Accession number	ORIGIN	TITLE
F04960	ORIGIN	Over-expression of a functionally active human GM2-activator

```

S-10-030-937-9 (1-193) x HSGM2APB (1-1093)
/dbc_type= mda;
/db xref="taxon:9606"

```

RESULT 8
SGM2AP

181 CCCCTGAGTCTCTCTCTGAAGGTGGATTAGTTTGGAGAGGAGGTGGCTGGCTCTGG 240
 Qy IleIysIleProCysThrAspTyrIleGlySerCysThrPheGluHisPheCysAspVal 114
 Db ATCAAGATCCCATGCACAGACTACATGGCAGCTGTACCTTGAACACTTCTGTGATGTG 300
 Qy LeuAspMetLeuIleProThrGlyGluProCysProGluProLeuArgThrTyrGlyLeu 134
 Db CTTGACATGTAATCTTACTTGGGAGCCCTGCCAGAGCCCTGCGTACCTATGGGGTT 360
 Qy ProCysHisCysProPheLeuGlyGlyThrTyrSerLeuProLysSerGluPheAlaVal 154
 Db CTTTGGCCACTGCTCCCTTCAAGAGAGGAACTTACTTACTTCCAGAGGGAATTCGTTGTG 420
 Qy ProAspLeuGluLeuProSerTyrPheLeuThrThrGlyAsnTyrArgIleGluSerValLeu 174
 Db CCTGACCTGGAGCTGCCAGTGGCTTACACCGGAACTACCGCATAGAGAGGCTCTGT 480
 Qy SerSerSerGlyLysArgLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db AGCAGCAGTGGGAAGGCTGGCTGCATCAAGATCGCTCTCTTAAAGGGCATA 537
 RESULT 9
 AB083313
 LOCUS Macaca fascicularis gm2a mRNA for ganglioside GM2 activator, 950 bp mRNA linear PRI 29-OCT-2003
 DEFINITION complete cds.
 AB083313
 ACCESSION AB083313.1 GI:23574732
 VERSION 2 (bases 1 to 950)
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Macaca fascicularis (crab-eating macaque)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae; Cercopitheinae; Macaca.
 REFERENCE 1
 AUTHORS Kusuda, J., Osada, N., Hida, M., Sugano, S. and Hashimoto, K.
 TITLE Isolation and characterization of cDNA for macaque neurological disease genes
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 950)
 AUTHORS Kusuda, J.
 TITLE Direct Submission
 JOURNAL Submitted (08-APR-2002) Jun Kusuda, National Institute of Infectious Diseases, Division of Genetic Resources, Toyama, Shinjuku, Tokyo 1628640, Japan (E-mail:jkusuda@nih.go.jp, URL:http://www.nih.go.jp, Tel:81-3-5285-1111(ex.2122), Fax:81-3-5285-1181)
 FEATURES
 source Location/Qualifiers
 1..950
 /organism="Macaca fascicularis"
 /mol_type="mRNA"
 /db_xref="taxon:9541"
 /clone="QcCE-17591"
 /tissue_type="brain cerebellum cortex"
 /clone_lib="macaque brain library QcCE"
 /note="vector:TOP10"
 1..950
 /gene="gm2a"
 66..638
 /gene="gm2a"
 /codon_start=1
 /product="ganglioside GM2 activator"
 /protein_id="BAC20592.1"
 /db_xref="GI:23574733"
 /translation="MQSLMQAPLVIALGLLFAAPQAHLKLGFSFMDNCDGKDPV
 IRSLTLPDPLILPNTVSVGTSVPLSPKLVLEKVALGIKIPCTDIYGS
 CTFEDSDVLDMLIPTGEGCPBPRTYGLPCHCFKGTYSLPKSEFVPHLELPSWL
 TTNRYIESILNSGRKLGCIKIAASLKV"
 ORIGIN
 Alignment Scores: 6.03e-78 Length: 950
 Pred. No.:

Score: 907.50 Matches: 172
 Percent Similarity: 93.26% Conservative: 8
 Best Local Similarity: 89.12% Mismatches: 10
 Query Match: 89.15% Indels: 3
 DB: 9 Gaps: 1
 US-10-030-937-9 (1-193) x AB083313 (1-950)
 Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuAlaThrPro 20
 Db 66 ATGCAGTCCCTGATGTCAGGCTCCCGTCTGATCGCCCTGGGCTTCTTTTCGGGCCCT 125
 Qy 21 AlaGlnAlaHisLeuLysLysProSerClnLeuSerSerPheSerThrAspAsnCysPhe 40
 Db 126 GCGCAAGCCCACTGAAAAAG-----CTTGGTAGCTTCTCTGGGATAACTGTGAT 176
 Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 177 GAAGGAAGGACCTCGGCTGATCAGAAAGCTGACTCTGGAGCCTGACCCCATCTCAT 236
 Qy 61 ProGlyAsnValThrLeuSerValIleGlySerThrSerValProLeuSerSerProLeu 80
 Db 237 CCTGGGAACGTGACTGTCTGAGTGTCTGGGAGCAGCAGTGTCTCTGAGTTCTCTCTG 296
 Qy 81 LysValAspLeuValLeuGlyLysGluValAlaGlyLeuTyrIleLysIleProCysThr 100
 Db 297 AAGGTGAGGTAGTCTGGAGAAAGAGTGGCTCTGGATCAAGATCCCATGCACA 356
 Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 357 GACTACATTGGCAGCTGTACCTTTGAAGACTCTCTGTGATGTACTGATGTATTCTCT 416
 Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 417 ACTGGGAGGGCTGCCAGAGCCCTGCTACTATGCGTCTCTGCTGCTGCTGCTGCTGCT 476
 Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuLeuLeuPro 160
 Db 477 AAAGAAGGAACCTACTACTCTGCCCAAGAGCGAATTCGTTGTGCTCTACCTGGAGTGC 536
 Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 Db 537 AGTTGGCTTACCCTGGAACTACCGCATAGAGAGCAATCTGAGCAACCGTGGGAAGCGT 596
 Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db 597 CTGGCTGCATCAAGATCGCTGCTCTCTTAAAGGGCGTA 635
 RESULT 10
 E12286
 LOCUS E12286 1983 bp DNA linear PAT 27-APR-1998
 DEFINITION cDNA encoding rat GM2 activator protein.
 ACCESSION E12286
 VERSION E12286.1 GI:3251120
 KEYWORDS JP 1996308582-A/1.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 1983)
 Takizawa, M. and Matsuo, N.
 RAT GM2 ACTIVATOR PROTEIN GENE
 TITLE Patent: JP 1996308582-A 1 26-NOV-1996;
 JOURNAL KAO CORP
 COMMENT OS Rattus sp.
 PN JP 1996308582-A/1
 PD 26-NOV-1996
 PF 23-MAY-1995 JP 1995123757
 PI TAKIZAWA MINORU, MATSUO NORI
 PC T12N15/09,C07H21/04//C07K14/47,C12N1/21,C12P21/02,(C12N1/21,
 PC C12R1:19),
 PC (C12P21/02,C12R1:19);

BC072474 LOCUS
DEFINITION Rattus norvegicus GM2 ganglioside activator protein, mRNA (CDNA clone MGC:91430 IMAGE:7100875), complete cds.
ACCESSION BC072474 GI:48735102
KEYWORDS MGC.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 2028) Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buettow,K.H., Schaefter,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,I., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,D., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huylk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Maman,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smialowski,D.E., Schnerch,A., Schein,J.E., Jones,S.O. and de Marra,M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) 12477932
REFERENCE 2 (bases 1 to 2028) Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk Email: cgabs-x@mail.nih.gov
Tissue Procurement: Howard Jacobs
CDNA Library Preparation: Express Genomics
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mcgonhri.nih.gov
Akhter,N., Ayelle,K., Beckstrom-Sternberg,S.M., Benjamin,B., Bialesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-I., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurganon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov Series: IRAK Plate: 175 Row: g Column: 14
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 27229315. Location/Qualifiers
1..2028 /organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="MGC:91430 IMAGE:7100875"

RESULT 14					
MUSGM2ACT					
LOCUS	MUSGM2ACT	1113 bp	mRNA	linear	ROD 02-FEB-1995
DEFINITION	Mouse GM2 activator protein mRNA,		complete cds.		

L19526
 L19526.1 GI:642678
 GM2 activator protein; activator protein; ganglioside GM2;
 lysosomal protein; lysosome.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1113)
 Bellachionia,G., Stirling,J.L., Orlacchio,A. and Beccari,T.
 Cloning and sequence analysis of a cDNA clone coding for the mouse
 GM2 activator protein
 Biochem. J. 294 (Pt 1), 227-230 (1993)
 93371367
 7689829
 On Feb 2, 1995 this sequence version replaced gi:309264.
 Original source text: Mus musculus cDNA to mRNA.

```

FEATURES             source
1..1113
Location/Qualifiers
    /organism="Mus musculus"
    /mol_type="mRNA"
    /db_xref="taxon:10090"
    /tissue_lib="lambda gt11 (from Clontech)"
14..595
CDS
    /codon_start=1
    /product="GM2 activator protein"
    /protein_id="AAA61929.1"
    /db_xref="GI:642679"

```

```

ORIGIN
Alignment Scores:
Pred. No.:          Length:      1113
Score:             738.50         134
Percent Similarity: 83.42%
Best Local Similarity: 69.43%
Query Match:       72.54%
DB:                10
US-10-030-937-9 (1-193) x MUSGM2ACT (1-1113)

Qy      3 SerLeuMetGlnAlaProLeuLeuLleAlaLeuGlyLeuLeuLeuAla-----Thr 19
      ::::      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      11 GCCATGACCGCTACCGCTCTGCTCTGCTGGGCTGTGCTGCGCAGGCTCCGTGCC 70

```



```

Db      165  GATGAAGGAAGACCTGCGAGTGATCAAAAGGCTACAGATCCAACTGACCCCATGTG 224
QY      60  ValProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerPro 79
Db      225  GTTCCTGGAGATGATGCTGCTGAGGCTGAGGGAAGACCAAGCGTTCCCTCAGTCTCCT 284
QY      80  LeuLysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCys 99
Db      285  CAGAAGGTGGAGCTCACCGTGGAGGAAGTGGCTTCTGGGTCAAGATTCTCTGT 344
QY      100 ThrAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIle 119
Db      345  GTAGAACAGCTAGGACGCTAGCTACGAGAACAATCTGTGACCTGATAGACGAATACATT 404
QY      120 ProThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysPro 139
Db      405  CCTCCTGGAGAGAGCTGCCAGAGCCCTGCACACCTACGGGCTGCCCTGCCACTGTCCC 464
QY      140 PheLysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeu 159
Db      465  TTCAAGGAGAGGTACTACTACTACCCAGCAGCAACTTCACAGTGCCTGACCTAGAGCTA 524
QY      160 ProSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLys 179
Db      525  CCGAGCTGGCTGAGCAGGCAACTCCGCATCCAGAGTATCTTTGAGCAGTGGTGGGAG 584
QY      180 ArgLeuGlyCysIleLysIleAlaAlaSerLeuLysGly 192
Db      585  CGCCTGGGCTGCATCAAGATTGCTGCTCTCTCAAGG3C 623

```

Search completed: November 18, 2004, 22:21:31
 Job time : 7864.24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 16:05:15 ; Search time 868.071 Seconds
(without alignments)
1167.114 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 1018

Sequence: 1 MQSLMQAPLLIALGILLATP.....LSSSGKRLGCIKIAASLKI 193

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Dlopp 6.0 , Delcxt 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2 1/USPTO.spool p/US10030937/runat 16112004 153013 2903/app query.fasta_1.789
-DB=N Geneseq 23Sep04 -QFMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937 @CGN 1 1240 @runat 16112004 153013 2903 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 23Sep04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2002bs:*
8: Geneseq2003as:*
9: Geneseq2003bs:*
10: Geneseq2003cs:*
11: Geneseq2003ds:*
12: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1005	98.7	1043	4 AAF54705	Aaf54705 Nucleotid
2	1005	98.7	1043	4 AAF54708	Aaf54708 Nucleotid
3	1005	98.7	2436	3 AAC55714	Aac55714 Human GM2
4	1005	98.7	2436	6 ABL63110	Ab163110 Breast ca
5	1005	98.7	2436	10 ADD71046	Add71046 Human GM2
6	1005	98.7	2436	11 ADN95859	Adn95859 Human BEC

7	1005	98.7	2478	6 ABV78068	Abv78068 Hypoxia-r
8	1005	98.7	2478	12 ADN03619	Adn03619 Antipsori
c	1005	98.7	2498	5 AAS81113	Aas81113 DNA encod
10	1000	98.2	953	12 ADQ17711	Adq17711 Human sof
11	1000	98.2	1935	10 ADB47402	Adb47402 Human cDN
12	1000	98.2	2384	6 ABK34915	Abk34915 Human cDN
13	1000	98.2	2471	5 AAS64907	Aas64907 DNA encod
14	1000	98.2	3988	12 ADQ22367	Adq22367 Human sof
15	774	76.0	579	4 AAF54730	Aaf54730 Nucleotid
16	761	74.8	579	4 AAF54698	Aaf54698 Nucleotid
17	740.5	72.7	1983	2 AAT61025	Aat61025 Rat GM2 a
18	740.5	72.7	1983	10 ADB52361	Adb52361 Primary r
19	544	53.4	546	5 AAS64904	Aas64904 DNA encod
20	535.5	52.6	1705	4 AAF54707	Aaf54707 Nucleotid
21	535.5	52.6	1706	4 AAF54701	Aaf54701 Nucleotid
22	535.5	52.6	1706	4 AAF54704	Aaf54704 Nucleotid
23	424	41.7	577	6 ABQ54556	Abq54556 Human ova
24	376.5	37.0	368	3 AAR42669	Aar42669 Human sec
c	354	34.8	475	4 AAI11589	Aai11589 Probe #15
c	354	34.8	475	4 ABA53278	Abas3278 Human foe
c	354	34.8	475	4 AAI32880	Aai32880 Probe #15
c	354	34.8	475	4 ABA42855	Abas42855 Human bre
c	354	34.8	475	4 ABA23052	Abas23052 Probe #15
c	354	34.8	475	4 AAK26980	Aak26980 Human bon
c	354	34.8	475	4 AAK01535	Aak01535 Human bra
c	354	34.8	475	4 ABS26569	Abs26569 Human liv
c	354	34.8	475	5 AAI01513	Aai01513 Probe #15
c	354	34.8	475	6 ABS01567	Abs01567 Human gen
c	354	34.8	475	6 ABQ60520	Abq60520 Human col
c	354	34.8	475	6 AAF54700	Aaf54700 Nucleotid
c	354	34.8	475	6 AAF54703	Aaf54703 Nucleotid
c	354	34.8	475	6 AAF54706	Aaf54706 Nucleotid
c	354	34.8	475	6 AAI20803	Aai20803 Probe #10
c	354	34.8	475	6 ABA65871	Abas65871 Human foe
c	354	34.8	475	6 AAI46033	Aai46033 Probe #14
c	354	34.8	475	6 ABA47983	Abas47983 Human bre
c	354	34.8	475	6 ABA32957	Abas32957 Probe #11
c	354	34.8	475	6 AAK40013	Aak40013 Human bon
c	354	34.8	475	6 AAK14282	Aak14282 Human bra

ALIGNMENTS

RESULT 1
AAF54705
ID AAF54705 standard; DNA; 1043 BP.
XX
AC AAF54705;
XX
DT 15-MAY-2001 (first entry)
XX
DE Nucleotide sequence of a human polymnucleotide sequence.

Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
neurological disease; auto-immune disease; multiple sclerosis; toxicity;
Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

Homo sapiens.

WO200105422-A2.

25-JAN-2001.

17-JUL-2000; 2000WO-FR002057.

15-JUL-1999; 99FR-00009372.

(INNR) BIOMERIEUX STELHYS.

Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

DR WPI; 2001-159475/16.
 XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 11; Page 181-182; 209pp; French.
 XX
 XX The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 1043 BP; 222 A; 303 C; 247 G; 271 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1-12e-95 Length: 1043
 Score: 1005.00 Matches: 191
 Percent Similarity: 98.96% Conservative: 0
 Best Local Similarity: 98.96% Mismatches: 2
 Query Match: 98.72% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x AAF54705 (1-1043)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuLeuLeuAlaThrPro 20
 DB 91 ATGCAGTCCCTGATGAGGCTCCCTCTGATGCCCTGGGCTTCTCGCGACCCCT 150
 QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 DB 151 GCGCAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTCCTGGGATACTGTGAT 210
 QY 41 GluGlyLysAspProAlaValLeuArgSerLeuThrLeuGluProAspProLeuVal 60
 DB 211 GAAGGGAAGACCTCGCGGTGATCAGAAGCTGACTCTGGAGCTGACCCCATCGTCT 270
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 DB 271 CCTGGAATGTGACCTCAGTGTCTGGGAGACACAGTGTCCCTGAGTTCCTCTG 330
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpLeuLysLeuProCysThr 100
 DB 331 AAGTGTGATTTAGTTTGGAGAGAGGTGGCTGGCTCTGGATCAGATCCCATGCACA 390
 QY 101 AspTrpLeuGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro 120
 DB 391 GACTACATTGGCAGCTACCTTTGAACACTCTGTGATGTGCTTGACATGTTAATTCCT 450
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 DB 451 ACTGGGAGCCCTCCAGAGACCCCTCGTACCTATGAGCTTCTGCTGCTGCTGCTGCT 510
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 DB 511 AAGAAGAGACCTACTCACTGCCAGAGCGAATTCGTGTGCTGCTGCTGCTGCTGCTGCT 570
 QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 DB 571 AGTTGGCTCACACCGGGAACCTACCGCATAGAGAGCGTCTCTGAGCAGCAGTGGAGCGT 630
 QY 181 LeuGlyCysIleLeuAlaAlaSerLeuLysGlyIle 193

DB 631 CTGGGCTGCATCAAGATCGCTGCTCTCTTAAGGGCATA 669
 RESULT 2
 AAF54708
 ID AAF54708 standard; DNA; 1043 BP.
 XX
 AC AAF54708;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of a human polynucleotide sequence.
 XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX (INMR) BIOMERIEUX STELHYS.
 PA
 XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 PI WPI; 2001-159475/16.
 XX
 DR Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 11; Page 183; 209pp; French.
 XX
 XX The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 1043 BP; 222 A; 303 C; 247 G; 271 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1-12e-95 Length: 1043
 Score: 1005.00 Matches: 191
 Percent Similarity: 98.96% Conservative: 0
 Best Local Similarity: 98.96% Mismatches: 2
 Query Match: 98.72% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x AAF54708 (1-1043)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuLeuLeuAlaThrPro 20
 DB 91 ATGCAGTCCCTGATGAGGCTCCCTCTGATGCCCTGGGCTTCTCGCGACCCCT 150

QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerThrAspAsnCysPhe 40
 DB 151 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTACTTTCTTGGGTAACCTGTGAT 210
 QY 41 GluGlyLysAspProAlaValAlaValLeuArgSerLeuThrLeuGluProAspProLeuVal 60
 DB 211 GAAGGGAAGGACCTGCGGTGATCAGAGCCCTGACTGTGGAGCCTGACCCCATGCTGTT 270
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 DB 271 CCTGGAATGTGACCTCAGTGTGCTGGGCGACACAGTGTCCCTGAGTTCTCTCTCTG 330
 QY 81 LysValAspLeuValLeuLysGluValAlaGlyLeuThrPheLeuLysLeuProCysThr 100
 DB 331 AAGGTGATTTAGTTTGGAGAGGAGTGGCTGGCTCTCGATCAAGATCCCATGCACA 390
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro 120
 DB 391 GACTACATTTGGCAGCTGTACCTTTGAACACACTCTGTGATGTGCTGACATGTTAATTCCT 450
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 DB 451 ACTGGGAGCCCTCCAGAGCCCTGCGTACCTATGGGCTTCTCTGCACTGTCTCTCTC 510
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 DB 511 AAAGAGGAACCTACTCTCTCCCAAGAGCAATTCGTTGGCTTGACCTGGAGCTGCC 570
 QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 DB 571 AGTTGGCTCACCCCGGACACTCCGCATAGAGAGGCTCTCTGAGCAGCAGTGGAGGGT 630
 QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 DB 631 CTGGGCTGCATCAAGATCGCTGCTCTCTCTAAAGGGGCATA 669
 RESULT 3
 AAC55714
 ID AAC55714 standard; cDNA; 2436 BP.
 XX
 AC AAC55714;
 XX
 DT 17-JAN-2001 (first entry)
 XX
 DE Human GM2 activator protein cDNA sequence from Genbank X62078.
 XX
 KW Human; differentially regulated gene; macrophage development; diagnosis;
 KW matrix metalloproteinase 19; MMP19; antiarthritic; antiinflammatory;
 KW destructive macrophage development inhibitor; arthritis;
 KW colorectal cancer; immune response; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO20005373-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 15-MAR-2000; 2000WO-US0006883.
 XX
 PR 15-MAR-1999; 99US-0124530P.
 XX
 PA (BOSB-) BOS BIOTECHNOLOGY INC.
 XX
 PI Murray R;
 XX
 XX WPI; 2000-628200/60.
 XX
 PT Screening drug candidates comprising adding a drug to a cell expressing an
 PT expression profile gene and determining the effect of the drug on the
 PT expression of the expression profile gene.
 XX
 PS Claim 1; Page; 99pp; English.
 XX

CC The present invention describes a method for screening drug candidates.
 CC The method comprises adding a drug to a cell that expresses an expression
 CC profile gene encoding a protein encoded by 5 sequences of defined base
 CC pairs as given in C55638, C55642, C55643, C55644 and C55653 or a sequence
 CC represented by Genbank accession number X92521, X62466, J04130, X62087
 CC and X76534 (or a fragment) and determining the effect of the drug on the
 CC expression of the expression profile gene. An inhibitor of matrix
 CC metalloproteinase 19 (MMP-19), preferably an antibody, is useful for
 CC treating destructive macrophage disorders (DMD) by inhibiting DM
 CC development in a cell of an individual having arthritis. Antibodies to
 CC MMP-19 are useful for localising a therapeutic moiety preferably
 CC cytotoxic agent or a radioisotope to colorectal cancer tissue. A
 CC composition comprising MMP-19 is useful for eliciting an immune response
 CC in an individual. C55635 to C55710 represent human differentially
 CC regulated genes of the invention. The present sequence represents the
 CC human GM2 activator protein cDNA sequence according to the Genbank
 CC accession number X62078. N.B. The present sequence is not given in the
 CC present specification, but it is specifically claimed by its Genbank
 CC accession number

XX SQ Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.58e-95 Length: 2436
 Score: 1005.00 Matches: 191
 Percent Similarity: 98.96% Conservative: 0
 Best Local Similarity: 98.96% Mismatches: 2
 Query Match: 98.72% Indels: 0
 DB: 3 Gaps: 0

US-10-030-937-9 (1-193) x AAC55714 (1-2436)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuAlaThrPro 20
 DB 59 ATGCAGTCCCTGATGAGGCTCCCTCTCTGATCGCCCTGGGCTTCTTCGCGACCCCT 118
 QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerThrAspAsnCysPhe 40
 DB 119 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTACTTTCTTGGGTAACCTGTGAT 178
 QY 41 GluGlyLysAspProAlaValAlaValArgSerLeuThrLeuGluProAspProLeuVal 60
 DB 179 GAAGGGAAGGACCTGCGGTGATCAGAGCCTGACTCTGGAGCCTGACCCCATGCTGTT 238
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 DB 239 CCTGGAATGTGACCTCAGTGTGCTGGCAGCACCAGTGTCCCTCTGAGTTCTCTCTG 298
 QY 81 LysValAspLeuValLeuLysGluValAlaGlyLeuThrPheLysIleProCysThr 100
 DB 299 AAGGTGATTTAGTTTGGAGAGGAGTGGCTGGCTCTGGATCAAGATCCCATGCACA 358
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro 120
 DB 359 GACTACATTTGGCAGCTGTACCTTTGAACACTCTGTGATGTGCTGATGTTAATTCCT 418
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 DB 419 ACTGGGAGCCCTGCCAGAGCCCTGCGTACCTATGGGCTTCTTGGCACTGTCCCTTC 478
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 DB 479 AAAGAGGAACCTACTCTCTCCCAAGAGCAATTCGTTGGCTTGACCTGAGCTGCC 538
 QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 DB 539 AGTTGGCTCACCCCGGACACTCCGCATAGAGGCTCTCTGAGCAGCAGTGGAGCGT 598
 QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 DB 599 CTGGGCTGCATCAAGATCGCTGCTCTCTCTAAAGGGGCATA 637

RESULT 4

ABL63110	PT	agent to be tested for anti-neoplastic activity, and determining a change
ID ABL63110 standard; DNA; 2436 BP.	PT	in expression of a gene of a signature gene set.
XX AC	XX	
XX ABL63110;	XX	
XX 15-MAY-2002 (first entry)	XX	
XX Breast cancer related gene sequence SEQ ID NO:1447.	XX	
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;	XX	
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;	XX	
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;	XX	
KW gene; ds.	XX	
XX Homo sapiens.	XX	
XX WO200194629-A2.	XX	
XX 13-DEC-2001.	XX	
XX 30-MAY-2001; 2001WO-US010838.	XX	
XX 05-JUN-2000; 2000US-0209473P.	XX	
XX 05-JUN-2000; 2000US-0209531P.	XX	
XX 18-SEP-2000; 2000US-0233133P.	XX	
XX 18-SEP-2000; 2000US-0233617P.	XX	
XX 20-SEP-2000; 2000US-0234009P.	XX	
XX 20-SEP-2000; 2000US-0234034P.	XX	
XX 20-SEP-2000; 2000US-0234052P.	XX	
XX 22-SEP-2000; 2000US-0234509P.	XX	
XX 22-SEP-2000; 2000US-0234567P.	XX	
XX 25-SEP-2000; 2000US-0234923P.	XX	
XX 25-SEP-2000; 2000US-0234924P.	XX	
XX 25-SEP-2000; 2000US-0235077P.	XX	
XX 25-SEP-2000; 2000US-0235082P.	XX	
XX 25-SEP-2000; 2000US-0235134P.	XX	
XX 25-SEP-2000; 2000US-0235280P.	XX	
XX 26-SEP-2000; 2000US-0235637P.	XX	
XX 26-SEP-2000; 2000US-0235638P.	XX	
XX 27-SEP-2000; 2000US-0235711P.	XX	
XX 27-SEP-2000; 2000US-0235720P.	XX	
XX 27-SEP-2000; 2000US-0235840P.	XX	
XX 27-SEP-2000; 2000US-0235863P.	XX	
XX 28-SEP-2000; 2000US-0236028P.	XX	
XX 28-SEP-2000; 2000US-0236032P.	XX	
XX 28-SEP-2000; 2000US-0236033P.	XX	
XX 28-SEP-2000; 2000US-0236109P.	XX	
XX 28-SEP-2000; 2000US-0236111P.	XX	
XX 29-SEP-2000; 2000US-0236842P.	XX	
XX 29-SEP-2000; 2000US-0236891P.	XX	
XX 02-OCT-2000; 2000US-0237172P.	XX	
XX 02-OCT-2000; 2000US-0237173P.	XX	
XX 02-OCT-2000; 2000US-0237278P.	XX	
XX 02-OCT-2000; 2000US-0237294P.	XX	
XX 02-OCT-2000; 2000US-0237295P.	XX	
XX 02-OCT-2000; 2000US-0237316P.	XX	
XX 03-OCT-2000; 2000US-0237425P.	XX	
XX 03-OCT-2000; 2000US-0237598P.	XX	
XX 03-OCT-2000; 2000US-0237604P.	XX	
XX 03-OCT-2000; 2000US-0237606P.	XX	
XX 03-OCT-2000; 2000US-0237608P.	XX	
XX 01-NOV-2000; 2000US-0244867P.	XX	
XX 01-NOV-2000; 2000US-0245084P.	XX	
XX (AVAL-) AVALON PHARM.	XX	
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;	XX	
PI Soppet DR, Weaver Z;	PI	
XX WPI; 2002-188264/24.	XX	
XX Screening for anti-neoplastic agent involves exposing cells to a chemical	XX	

PT	agent to be tested for anti-neoplastic activity, and determining a change
PT	in expression of a gene of a signature gene set.
XX PS	Claim 1; SEQ ID NO 1447; 44pp; English.
XX	
XX	The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's tumour
XX	
XX	Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;
XX	
XX	Alignment Scores:
XX	Pred. No.: 3,588-95 Length: 2436
XX	Score: 1005.00 Matches: 191
XX	Percent Similarity: 98.96% Conservative: 0
XX	Best Local Similarity: 98.96% Mismatches: 2
XX	Query Match: 98.72% Indels: 0
XX	DB: 6 Gaps: 0
XX	
XX	US-10-030-937-9 (1-193) x ABL63110 (1-2436)
QY	1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db	59 ATGCAGTCCCTGATGCAGGCTCCCTCTGATCGCTGGCTTGTCTTCGCGAGCCCT 118
QY	21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerPheSerPheSerPhe 40
Db	119 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTTCCTGGGATACTGTGAT 178
QY	41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db	179 GAAGGGAAGACCTCGGTGATCAGAGCTGACTCTGGAGCTGACCCCATCGTCTGTT 238
QY	61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db	239 CCTGGAATGTCAGCTCAGTGTGGGCGAGCAGTGTCCCTGAGTTCCTCTCTG 298
QY	81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrTrpIleLysIleProCysThr 100
Db	299 AAGGTGGATTAGTTTTGGAGAAGAGGTGGCTCTGGATCAAGATCCCATGCACACA 358
QY	101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db	359 GACTACATTGGCAGCTGTACTTTTGAACCTTCTGTGATGTGCTTGACATGTTAATTCCT 418
QY	121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db	419 ACTGGGAGCCCTGCCAGAGCCCTCGCTACCTATGGGCTTCTTCCCACTGTCCCTTC 478
QY	141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
Db	479 AAAGAAGGAACCTACTACTGCCCAAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 538
QY	161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
Db	539 AGTTGGCTCACCACCGGGAACCTACCGCATAGAGAGCGTCTTGAGCAGCAGTGGAGCGT 598
QY	181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193

Db 599 CTGGCTGCATCAAGATCGCTGCTCTCTTAAGGGCATA 637

RESULT 5
ADD71046
ID ADD71046 standard; DNA; 2436 BP.
XX AC ADD71046;
XX DT 15-JAN-2004 (first entry)
XX DE Human GM2 ganglioside activated protein gene SEQ ID NO:50.
XX KW liver cancer; chronic hepatitis; cirrhosis; liver disease; hepatotropic;
XX KW cytotatic; gene therapy; human; gene; ds.
XX OS Homo sapiens.
XX FN WO2003061564-A2.
XX PD 31-JUL-2003.
XX PF 20-DEC-2002; 2002WO-US040718.
XX PR 21-DEC-2001; 2001US-0341815P.
XX PR 31-DEC-2001; 2001US-0343185P.
XX PA (GENE-) GENE LOGIC INC.
XX PA (LGBI-) LG BIOMEDICAL INST.
XX Koh SS, Liu Q, Chung H, Zeng W, Lee B, Yeramilli S, Song SY;
XX WPI; 2003-663343/62.
XX
XX Diagnosing liver cancer cells, useful for treating liver cancer
XX associated with chronic hepatitis or cirrhosis comprises detecting the
XX level of expression in a tissue sample of one or more genes associated
XX with cancerous liver tissues.
XX
XX Claim 1; SEQ ID NO 50; 176pp; English.
XX
XX The present invention describes a method for diagnosing liver cancer
XX cells comprising detecting the level of expression in a tissue sample of
XX one or more genes given in the specification (see ADD70997 to ADD71105),
XX where differential expression of the genes is indicative of liver cancer.
XX Also described: (1) detecting the progression of liver cancer in a
XX patient; (2) monitoring the treatment of a patient with liver cancer; (3)
XX treating a patient with liver cancer; (4) typing a liver disease in a
XX patient; (5) detecting the presence or progression of liver cancer in a
XX patient with chronic hepatitis or cirrhosis; (6) differentiating liver
XX cancer related to chronic hepatitis from liver cancer related to
XX cirrhosis; (7) screening for an agent capable of modulating the onset or
XX progression of liver cancer; (8) a composition comprising at least two
XX oligonucleotides comprising a sequence that specifically hybridizes to
XX any of the genes; (9) a solid support comprising the at least two
XX oligonucleotides; (10) a computer system comprising a database containing
XX information identifying the level in liver tissue of a set of genes; (11)
XX a method for using the computer system to present information identifying
XX the expression level in tissue or cell of any of the genes; and (12) a
XX therapeutic agent for slowing or halting the progression of liver cancer.
XX The methods are useful for treating liver cancer associated with chronic
XX hepatitis or cirrhosis. The present sequence represents a specifically
XX claimed human gene sequence which is used in the exemplification of the
XX present invention.
XX
XX Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3.58e-95 Length: 2436
Score: 1005.00 Matches: 191
Percent Similarity: 98.96% Conservatives: 0
Best Local Similarity: 98.96% Mismatches: 2
Query Match: 98.72% Indels: 0
DB: 10 Gaps: 0

US-10-030-937-9 (1-193) x ADD71046 (1-2436)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
DB 59 ATGAGTCCCTGATGACAGGCTCCCTCTCTGATCGCCCTGGGCTTCTTCTCGGACCCCT 118
QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTyrAspAsnCysPhe 40
DB 119 GCGCAAGCCACCTGAAAAAGCCATCCAGCTCAGTAGCTTTCTCTGGGATAACTGTGAT 178
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
DB 179 GAAGGGAAGGACCCCTGCGGTGATCAGAAAGCTGACTCTGGAGCCCTGACCCCATCGTCT 238
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
DB 239 CCTGGAAATGTGACCCCTCAGTGTGCTGGGACAGCAGTGTCCCTCCTGAGTTCTCTCTG 298
QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPheIleLysIleProCysThr 100
DB 299 AAGGTGGATTTAGTTTGGAGAGAGGTGGCTGCTGATCAAGATCCCATGACACA 358
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
DB 359 GACTACATTTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTTGACATGTTAATCT 418
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
DB 419 ACTGGGAGCCCTGCCAGAGCCCTCGGTACCTATGGGCTTCTTGCACATGTCCTTC 478
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
DB 479 AAGAAGGAACCTACTACTCCTCCCAAGAGCGAATTCGTTGTGCTGACCTGAGCTGCC 538
QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyIlyArg 180
DB 539 AGTTGGCTCACCCACCCGGAACCTACCCATAGAGAGCGTCTCTGAGCAGCAGTGGGAAGCGT 598
QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
DB 599 CTGGCTGCATCAAGATCGCTGCTCTCTTAAGGGCATA 637

RESULT 6
ADN95859
ID ADN95859 standard; DNA; 2436 BP.
XX AC ADN95859;
XX DT 01-JUL-2004 (first entry)
XX DE Human BEC/LEC-related gene sequence SeqID783.
XX KW growth; differentiation; blood endothelial cell; BEC;
XX KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
XX KW lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytotatic;
XX KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
XX KW inflammatory disease; cancer metastasis; lymphatic system; gene; ds;
XX human.
XX OS Homo sapiens.
XX FN WO2003080640-A1.
XX PD 02-OCT-2003.
XX PF 07-MAR-2003; 2003WO-US006900.
XX PR 07-MAR-2002; 2002US-0363019P.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PA (LICN) LICENTIA LTD.
XX

PI Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen P, Saharinen J;
XX WPI; 2003-876899/81.
XX P-PSDB; ADN95858.
XX
PS Example 1; SEQ ID NO 783; 176pp; English.
XX
XX This invention relates to a method of differentially modulating the
CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
CC endothelial cells (LEC) comprises contacting endothelial cells with a
CC composition comprising an agent that differentially modulates blood or
CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
CC identifying a human subject with lymphoedema and with a mutation in at
CC least one allele of a gene encoding a LEC protein, where the mutation
CC correlates with lymphoedema in human subjects, and with the proviso that
CC the LEC protein is not VEGFR-3; and administering to the subject a
CC composition comprising a lymphatic growth agent selected from VEGF-C or
CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
CC the development of compounds with an antiangiogenic, cytostatic,
CC vasotropic or antiinflammatory activity or for gene therapy. The method
CC is useful in modulating the growth or differentiation of blood
CC endothelial cells or lymphatic endothelial cells, in treating hereditary
CC lymphoedema, in screening for an endothelial cell disorder or
CC predisposition to the disorder or in monitoring the efficacy or toxicity
CC of a drug on endothelial cells. The agent is useful in manufacturing a
CC medicament for the differential modulation of blood vessel endothelial
CC cell or lymphatic vessel endothelial cell growth or differentiation. The
CC lymphatic growth agent may also be used in manufacturing a medicament for
CC the treatment of hereditary lymphoedema resulting from a mutation in a
CC LEC gene or of other diseases involving the lymphatic vessels, such as
CC various inflammatory diseases and cancer metastasis via the lymphatic
CC system. The present sequence is that of a human LEC/BEC differentially
CC expressed gene which is related to the method of the invention. Note: This
CC sequence does not appear in the specification but was obtained by the
CC index using the source data given in table 14 of the specification.
XX
SQ Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,59e-95 Length: 2436
Score: 1005.00 Matches: 191
Percent Similarity: 98.96% Conservative: 0
Best Local Similarity: 98.96% Mismatches: 2
Query Match: 98.72% Indels: 0
DB: 11 Gaps: 0

US-10-030-937-9 (1-193) x ADN95859 (1-2436)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db 59 ATGCAGTCCCTGATCAGCGCTCCCTCCTGATCGCCCTGGGCTTGCTTCGCGACCCCT 118
QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
Db 119 GCGCAAGCCACCTGAAAAGGCCATCCAGCTCAGTAGCTTTCTCGGATAACTGTGAT 178
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 179 GAAGGGAAGGACCTTCGCTGATCAGAGAGCCCTGACTCGAGGCTGACCCCATCGTCGPT 238
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 239 CCTGGAATGTGACCTCAGTGTGTTGGGCAGCACAGTGTCCCTGAGTTCTCTCTCTG 298
QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
Db 299 AAGGTGGATTAGTTTTTGCAGAGAGGAGGTGGCTGCTCGATCAAGATCCCATGCACA 358
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 359 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTTGACATGTTAATCTCT 418
QY 121 ThrGlyGluProCysProGluProLeuA:gtThrTyrGlyLeuProCysHisCysProPhe 140

CC retinopathy, neonatal stress, pre-eclampsia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss
 XX Sequence 2478 BP; 588 A; 638 C; 586 G; 666 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3.66e-95 Length: 2478
 Score: 1005.00 Matches: 191
 Percent Similarity: 98.96% Conservatives: 0
 Best Local Similarity: 98.96% Mismatches: 2
 Query Match: 98.72% Indels: 0
 DB: 6 Gaps: 0

US-10-030-937-9 (1-193) x ABV78068 (1-2478)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 Db 96 ATGCAGTCCCTGATGCAGGCTCCCTCTCTGATCGCCCTGGGCTTCTTCTCGGACCCCT 155
 QY 21 AlaGlnAlaHisLeuLysLeuLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 Db 156 GCGAAGGCCACCTGAAAGACCATCCAGCTCAGTAGCTTTCTCTGGGATAAATCTGTGAT 215
 QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 216 GAAGGGAAGGACCTCGCGTGTATCAGAGCCCTGACTCTGGAGCCTGACCCCATCTGCTGT 275
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 Db 276 CCTGGAATGTGACCTCAGTGTCTGGGAGCAGCAGTGTCCCTGAGTCTCTCTCTG 335
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPheLeuLysLeuProCysThr 100
 Db 336 AAGGTGGATTAGTTTGGAGAGAGGTGGTGGCTCTGTGATCAAGATCCCATGACACA 395
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 396 GACTACATTTGGCAGCTGACCTTGAACACTTCTGTGATGTGCTTACATGTTAATTCCT 455
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 456 ACTGGGAGCCCTGCCAGAGCCCTGCTATGGCTTCTTCTGCTGCTTCTCTCTCTCT 515
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 Db 516 AAAGAGGAACCTACTACTCTGCCAAGAGCGAATTCGTTGTGCTGACCTGGAGTGCC 575
 QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 Db 576 AGTTGGCTCACCCGGAACCTACCGCATAGAGAGCGTCTCTGAGCAGCAGTGGAGCGGT 635
 QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db 636 CTGGGCTGCATCAAGATCGCTGCTCTCTAAAGGGCATA 674

RESULT 8

ADN03619
 ID ADN03619 standard; cDNA; 2478 BP.

AC ADN03619;

DT 01-JUL-2004 (first entry)

DE Antipsoriatic cDNA sequence #7.

XX ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.

XX Homo sapiens.

XX WO2004028479-A2.

XX 08-APR-2004.

XX

PF 25-SEP-2003; 2003WO-US030907.
 XX 25-SEP-2002; 2002US-0414006P.
 PR (GETH) GENENTECH INC.
 PA Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
 PI Wu TD;
 PI WPI; 2004-305105/28.
 XX P-PSDB; ADN03620.
 DR New PRO nucleic acid or polypeptide, useful for preparing a
 PT pharmaceutical composition for diagnosing or treating psoriasis in a
 PT mammal.
 XX Claim 1; SEQ ID NO 13; 3069pp; English.
 PS The invention relates to novel polynucleotide and polypeptides for
 CC treating psoriasis or a sequence having at least 80% identity to the
 CC above sequences. The nucleic acid is useful for preparing a composition
 CC for diagnosing or treating psoriasis in a mammal. This sequence
 CC corresponds to one of the polynucleotides of the invention.
 XX SQ Sequence 2478 BP; 588 A; 638 C; 586 G; 666 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.66e-95 Length: 2478
 Score: 1005.00 Matches: 191
 Percent Similarity: 98.96% Conservatives: 0
 Best Local Similarity: 98.96% Mismatches: 2
 Query Match: 98.72% Indels: 0
 DB: 12 Gaps: 0

US-10-030-937-9 (1-193) x ADN03619 (1-2478)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 Db 96 ATGCAGTCCCTGATGCAGGCTCCCTCTCTGATCGCCCTGGGCTTCTTCTCGGACCCCT 155
 QY 21 AlaGlnAlaHisLeuLysLeuLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 Db 156 GCGAAGGCCACCTGAAAGACCATCCAGCTCAGTAGCTTTCTCTGGGATAAATCTGTGAT 215
 QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 216 GAAGGGAAGGACCTCGCGTGTATCAGAGCCCTGACTCTGGAGCCTGACCCCATCTGCTGT 275
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 Db 276 CCTGGAATGTGACCTCAGTGTCTGGGAGCAGCAGTGTCCCTGAGTCTCTCTCTCTG 335
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPheLeuLysLeuProCysThr 100
 Db 336 AAGGTGGATTAGTTTGGAGAGAGGTGGTGGCTCTGTGATCAAGATCCCATGACACA 395
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 396 GACTACATTTGGCAGCTGACCTTGAACACTTCTGTGATGTGCTTACATGTTAATTCCT 455
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 456 ACTGGGAGCCCTGCCAGAGCCCTGCTATGGCTTCTTCTGCTGCTTCTCTCTCTCT 515
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 Db 516 AAAGAGGAACCTACTACTCTGCCAAGAGCGAATTCGTTGTGCTGACCTGGAGTGCC 575
 QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 Db 576 AGTTGGCTCACCCGGAACCTACCGCATAGAGAGCGTCTCTGAGCAGCAGTGGAGCGGT 635
 QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193

PT and comparing the gene expression, also useful in treating soft tissue
 XX sarcoma.
 PS Example 2; SEQ ID NO 528; 210pp; English.
 XX
 CC The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytosolic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX
 SQ Sequence 953 BP; 201 A; 286 C; 231 G; 235 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3.3e-95 Length: 953
 Score: 1000.00 Matches: 190
 Percent Similarity: 98.45% Conservative: 0
 Best Local Similarity: 98.45% Mismatches: 3
 Query Match: 98.23% Indels: 0
 DB: 12 Gaps: 0

US-10-030-937-9 (1-193) x ABQ17711 (1-953)
 QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 Db 91 ATGCAGTCCCTGATGAGGCTCCCTCTGATCGCCCTGGGCTTGGCTTCGCGGCCCT 150
 QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerThrAspAsnCysPhe 40
 Db 151 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTCTGGGATACTGTGAT 210
 QY 41 GluGlyLysAspProAlaValAlaHisArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 211 GAAGGGAAGACCCCTGGGGTATCAGAGGCTGACTCTGGAGCTGACCCCATCGTGT 270
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 Db 271 CCTGGAATGTGACCTTCAGTGTCTGGGAGACCCAGTGTCCCTCGAGTTCCTCTG 330
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPilleLysIleProCysThr 100
 Db 331 AAGGTGGATTAGTTTGGAGAGGAGTGGCTCTGGATCAAGATCCCATGTCACA 390
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro 120
 Db 391 GACTACATTTGGAGCTGTACCTTTGAACACTTCTGTGATGTGCTGACATGTTAATTCCT 450
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 451 ACTGGGAGCCCTGCGCAGAGCCCTCTGGTACCTATGGGCTTCTTGGCCACTGTCCCTTC 510
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 Db 511 AAAGAAGGACCTTACTCAGTCCCAAGAGCGAATTCGTTGTGCTGACCTGGAGTGGCC 570
 QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysArg 180
 Db 571 AGTTGGCTACACCGGAACTACCGCATAGAGAGCGTCTGAGCAGCAGTGGGAGCGT 630
 QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db 631 CTGGGCTGCATCAAGATCGCTGCTCTCTAAAGGGGATA 669

RESULT 11
 ADB47402

ID ADB47402 standard; cDNA; 1935 BP.
 XX
 AC ADB47402;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Human cDNA upregulated in dendritic cells SEQ ID NO 102.
 XX
 KW ss; gene; human; dendritic cells; high throughput; cancer;
 KW infectious disease; autoimmune disease; allergy;
 KW graft versus host disease; vaccine enhancing; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN US2003134283-A1.
 XX
 PD 17-JUL-2003.
 XX
 PF 03-OCT-2001; 2001US-00971392.
 XX
 PR 03-OCT-2000; 2000US-0237652P.
 XX
 PA (PETE/) PETERSON D P.
 PA (PEAR/) PEARSON C I.
 PA (COCK/) COCKS B G.
 XX
 PI Peterson DP, Pearson CI, Cocks BG;
 XX
 DR WPI; 2003-662509/62.
 XX
 PT New combination comprises cDNAs that are differentially expressed in
 PT dendritic cells useful for preparing a composition for diagnosing or
 PT treating cancer, infectious disease, autoimmunity, allergy or graft
 PT versus host disease.
 XX
 PS Claim 1; SEQ ID NO 102; 28pp; English.
 XX
 CC The invention relates to a combination comprising cDNAs that are
 CC differentially expressed in dendritic cells (DC). Also included is a high
 CC throughput method for detecting differential expression of one or more
 CC cDNAs in a sample containing nucleic acids. The combination is useful for
 CC preparing a composition for diagnosing, treating and monitoring the
 CC treatment of cancer, infectious disease, autoimmunity, allergy or graft
 CC versus host disease, or for enhancing a vaccine. The present sequence
 CC represents a human cDNA upregulated in dendritic cells. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?DocID=20030134283.
 XX
 SQ Sequence 1935 BP; 459 A; 507 C; 460 G; 509 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.74e-95 Length: 1935
 Score: 1000.00 Matches: 190
 Percent Similarity: 98.45% Conservative: 0
 Best Local Similarity: 98.45% Mismatches: 3
 Query Match: 98.23% Indels: 0
 DB: 10 Gaps: 0

US-10-030-937-9 (1-193) x ADB47402 (1-1935)
 QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 Db 102 ATGCAGTCCCTGATGAGGCTCCCTCTGATCGCCCTGGGCTTGGCTTCGCGGCCCT 161
 QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerThrAspAsnCysPhe 40
 Db 162 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTCTGGGATACTGTGAT 221
 QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 222 GAAGGGAAGGACCCCTGCGGTGATCAGAGCTGACTCTGGAGCTGACCCCATCGTGT 281

QY	61	ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu	80
Db	282	CCTGGAAATGTACCCCTAGTCTCGTGGCAGCACCAGTCCCTCGATTCTCTCTG	341
QY	81	LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysThrProCysThr	100
Db	342	AAGGTGGATTTAGTTTGGAGAGAGAGGTGGCTCTGGATCAAGATCCATGCACA	401
QY	101	AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro	120
Db	402	GACTACATTTGGAGCTGTACCTTTGAACATTTCTGTGATGTCTTGACATGTTAAATCT	461
QY	121	ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe	140
Db	462	ACTGGGAGCCCTGCGCACAGCCCTCGCTACTATGGGCTTCTTGGCCACTGTCCCTTC	521
QY	141	LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro	160
Db	522	AAAGAGGAACTACTCTACCTGCCCAAGAGCGAATTCGTTGCTGACCTGGAGTGGCC	581
QY	161	SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysArg	180
Db	582	AGTTGGCTCACCCACCGGAACTACCGCATAGAGAGGTCTCTGAGCAGCAGTGGAGCGT	641
QY	181	LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle	193
Db	642	CTGGGCTGCATCAAGATCGTGCCTCTCTAAAGGGCATA	680
RESULT 12			
ID	ABK34915	standard; cDNA; 2384 BP.	
AC	AC		
XX	ABK34915;		
XX			
DT	08-MAY-2002	(first entry)	
XX			
DE	Human cDNA encoding secreted protein #53.		
KW	Human; secreted protein; gene; ss; nutritional supplement; haemophilia;		
KW	viral infection; bacterial infection; fungal infection; diabetes; asthma;		
KW	autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;		
KW	autoimmune thyroiditis; allergic reaction; neurodegenerative disease;		
KW	Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;		
KW	coagulation disorder; inflammatory disorder; Crohn's disease; incision;		
KW	tissue regeneration; wound healing; burn; haematopoiesis;		
KW	myeloid cell deficiency; lymphoid cell deficiency.		
OS	Homo sapiens.		
XX			
XX	WO200177288-A2.		
XX			
PD	18-OCT-2001.		
XX			
PF	29-MAR-2001; 2001WO-US010224.		
XX			
PR	06-APR-2000; 2000US-0195582P.		
XX			
PA	(GEMY) GENETICS INST INC.		
XX			
PI	Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;		
PI	Gulukota K, Graham JR;		
XX			
DR	WPI; 2002-179321/23.		
XX			
PT	Five hundred and ninety two polynucleotides derived from a variety of		
PT	human tissue sources which encode secreted proteins, useful for treating		
PT	immune deficiencies and disorders such as autoimmune disorders.		
XX			
PS	Claim 1; Page 95-96; 372pp; English.		
XX			
CC	The invention relates to 592 polynucleotides which have been derived from		
CC	a variety of human tissue sources and which encode novel secreted		
CC	proteins. The polynucleotides can be used as probes for the		

CC	identification and isolation of full length cDNA and genomic DNA. The		
CC	polynucleotides and proteins can also be used as nutritional supplements.		
CC	The proteins are useful in the treatment of various immune deficiencies.		
CC	and disorders such as viral infections, bacterial infections, fungal		
CC	infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple		
CC	sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions		
CC	and conditions (e.g. asthma). They are also useful for treating		
CC	neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's		
CC	disease), liver fibrosis, coagulation disorders (e.g. haemophilia),		
CC	inflammatory disorders (e.g. Crohn's disease) and tumours. They are also		
CC	useful for tissue regeneration, for wound healing and in the treatment of		
CC	burns, incisions and ulcers. The proteins are also useful for regulating		
CC	haematopoiesis and for treating myeloid or lymphoid cell deficiencies.		
CC	Sequences ABK34863-ABK35454 represent polynucleotides of the invention		
XX			
SQ	Sequence 2384 BP; 550 A; 618 C; 571 G; 645 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	1.16e-94	Length:	2384
Score:	1000.00	Matches:	190
Percent Similarity:	98.45%	Conservative:	0
Best Local Similarity:	98.45%	Mismatches:	3
Query Match:	98.23%	Indels:	0
DB:	6	Gaps:	0
US-10-030-937-9 (1-193) x ABK34915 (1-2384)			
QY	1	MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuAlaThrPro	20
Db	13	ATGCAGTCCCTGATCAGGCTCCCTCTGATCGCCCTGGGCTGTTCTCGGGCCCT	72
QY	21	AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe	40
Db	73	GCGCAAGCCACCTGAAAGGCCATCCAGCTCAGTAGCTTTTCTGGGATACTGTGAT	132
QY	41	GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal	60
Db	133	GAAGGGAAGGACCTCGGGTGATCAGAAGCCTGACTCTGGAGCCTGACCCATCGCTGT	192
QY	61	ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu	80
Db	193	CCTGGAAATGTGACCCCTCAGTGTCTGGGAGCAGCAGTGTCCCTCTGAGTTCCTCTG	252
QY	81	LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysThrProCysThr	100
Db	253	AAGGTGGATTTAGTTTGGAGAGAGGTGGCTGGGCTCTGGATCAGATCCATGCACA	312
QY	101	AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro	120
Db	313	GACTACATTTGGCAGCTGACTTTGAACACTTCTGTGATGTCTGACATGTTAAATCTCT	372
QY	121	ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe	140
Db	373	ACTGGGAGCCCTGCCAGAGCCCTCGTACCTATGGGCTTCTTGGCCACTGTCCCTTC	432
QY	141	LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro	160
Db	433	AAAGAGGAACTACTCTACTGCCCAAGAGCGAATTCGTTGTGCTGACCTGGAGTGGCC	492
QY	161	SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg	180
Db	493	AGTTGGCTCACCCAGGGAATACCGCATAGAGAGCGTCTCTGAGCAGCAGTGGGAGCGT	552
QY	181	LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle	193
Db	553	CTGGGCTGCATCAAGATCGCTGCCTCTCTAAAGGGCATA	591
RESULT 13			
ID	AAS64907	standard; cDNA; 2471 BP.	
XX	AAS64907;		
AC			
XX			

DT 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #711.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX WO200175067-A2.
 FN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US008631.
 PF 31-MAR-2000; 2000US-00540217.
 FR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 XX P-ESDB; ABG00720.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 711; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2471 BP; 578 A; 639 C; 583 G; 671 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.22e-94 Length: 2471
 Score: 1000.00 Matches: 190
 Percent Similarity: 98.45% Conservative: 0
 Best Local Similarity: 98.45% Mismatches: 3
 Query Match: 98.23% Indels: 0
 DB: 5 Gaps: 0
 US-10-030-937-9 (1-193) x AAS64907 (1-2471)
 QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 DB 93 ATGCAGTCCCTGATGCGAGCTCCCTCTGATCGCCCTGGGCTTGGCTTCGCGGCCCT 152
 QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 DB 153 GCGCAAGCCACCTGAAAGAACCATCCAGCTCAGTAGCTTTCTCTGGGTAATCTGTGAT 212

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 DB 213 GAAGGGAAGGACCCCTGCGGTGATCAGAAGCCTGAGCTCTGGAGCCTGACCCCATCGTGT 272
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 DB 273 CCTGGAATGTGACCCCTCAGTGTGCTGGGCGACACAGTGTCCCTCGAGTCTCTCTCTG 332
 QY 81 LysValAspLeuValLeuGlyLysGluValAlaGlyLeuTrpIleLysLysProCysThr 100
 DB 333 AAGGTGGATTAGTTTGGAGAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGCACA 392
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro 120
 DB 393 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTTGACATGTTAATCTCT 452
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 DB 453 ACTGGGAGCCCTGCCAGAGCCCTGCGTACCTATGGGCTTCCCTGCCACTGTCCCTTC 512
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 DB 513 AAGAGGAGAACTACTACTCTGCCCAAGAGCGAATTCGTTGGCTGACCTGGAGCTGCC 572
 QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 DB 573 AGTTGGCTCACACCGGGAACCTACCGCATAGAGCGTCTCTGAGCAGTGGGAAGCGT 632
 QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 DB 633 CTGGGCTGCATCAAGATCGCTGCCTCTCTAAAGGGCATA 671

RESULT 14
 ADQ22367
 ID ADQ22367 standard; DNA; 3988 BP.

AC ADQ22367;
 XX 26-AUG-2004 (first entry)
 DT Human soft tissue sarcoma-upregulated DNA - SEQ ID 5187.
 DE soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
 XX ds.
 KW Homo sapiens.
 OS WO2004048938-A2.
 FN 10-JUN-2004.
 PD 26-NOV-2003; 2003WO-US038193.
 PF 26-NOV-2002; 2002US-0429739P.
 PR (PROT-) PROTEIN DESIGN LABS INC.
 PA Aziz N, Ginsburg WM, Zlotnik A;
 FI WPI; 2004-441208/41.
 XX Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.

Example 2; SEQ ID NO 5187; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the

CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytotatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX
 SQ Sequence 3988 BP; 1094 A; 927 C; 935 G; 986 T; 0 U; 46 Other;

Alignment Scores:
 Pred. No.: 2,36e-94 Length: 3988
 Score: 1000.00 Matches: 190
 Percent Similarity: 98.45% Conservative: 0
 Best Local Similarity: 98.45% Mismatches: 3
 Query Match: 98.23% Indels: 0
 DB: 12 Gaps: 0

US-10-030-937-9 (1-193) x ADQ22367 (1-3988)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 Db 96 ATGCAGTCCCTGATGAGGCTCCCTCTGATGCGCTGGCTTGTCTCGGGCCCT 155
 QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 Db 156 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTCTCTGGGATACTGTGAT 215
 QY 41 GluGlyAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 216 GAAGGGAAGACCTCGCGTGATCAGAAGCTTGACTCTGGAGCTGACCCATCGTGT 275
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 Db 276 CTGGAAATGTGACCTCAGTCTGCTGGGAGCAGCAGTGTCCCTGAGTCTCCTCTG 335
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
 Db 336 AAGGTGATTTAGTTTGGAGAAGAGGTGGCTGGCTCTGGATCAAGATCCCATGCA 395
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 396 GACTACATTTGGAGCTTACCTTTGAACATCTCTGTGATGCTTGATCTTAATTCCT 455
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 456 ACTGGGAGCCCTGCCCAGAGCCCTCGGTACCTATGCGCTTCTCTGCCACTGTCCCTTC 515
 QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 Db 516 AAAGAGGAACCTACTCACTGCCCAAGAGCGAATTGCTGTGCTGACCTGGAGCTGCC 575
 QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleLeuSerValLeuSerSerGlyLysArg 180
 Db 576 AGTTGGCTACCAACCGGAACCTACCGCATAGAGAGCGTCTCTGACGACGAGTGGAGCGT 635
 QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db 636 CTGGCTGTCATCAAGATCGTCTCTCTAAAGGGCATA 674

RESULT 15
 AAF54730
 ID AAF54730 standard; DNA; 579 BP.
 XX
 AC AAF54730;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of a human polynucleotide sequence.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;

KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthriti; lupus erythematosus; gene therapy; ss.
 XX Homo sapiens.
 XX WO200105422-A2.
 PN 25-JAN-2001.
 XX 17-JUL-2000; 2000WO-FR002057.
 XX 15-JUL-1999; 99FR-00009372.
 PR (INMR) BIOMERIEUX STELHYS.
 PA Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 PI WPI; 2001-159475/16.
 DR Detecting, preventing and treating degenerative, neurological and
 XX autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX Claim 11; Page 208; 209pp; French.
 PS The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthriti and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 579 BP; 83 A; 66 C; 89 G; 82 T; 0 U; 259 Other;
 Alignment Scores:
 Pred. No.: 9.05e-72 Length: 579
 Score: 774.00 Matches: 141
 Percent Similarity: 73.06% Conservative: 0
 Best Local Similarity: 73.06% Mismatches: 52
 Query Match: 76.03% Indels: 0
 DB: 4 Gaps: 0
 US-10-030-937-9 (1-193) x AAF54730 (1-579)
 QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 Db 1 ATGCARWSYNTATGCGAGCCNCCNYNTYNTATGCTNTGNGNYNTYNTYNGCNACCCN 60
 QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 Db 61 GCNCARGCNCAYYTNAAARCCNWSNRCARYTNWSNNTYWSNTGGAYAYTGT 120
 QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 121 GARGGNAARGAYCCNGCNGTNGTHMGWSNNTYNTACNTNGARCCNGAYCCNATHGTNGTN 180
 QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 Db 181 CCNGGNAAYGTNACNTYNTWSNGTNGTNGNWSNACNWSNNGTNCNCCNYTNWSNCCNYTN 240
 QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100


```

Db      241 AARGTNGAYTTGNTNGARARGTNGCNGGNTNTGGATHAARATHCCNTGYACN 300
QY      101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db      301 GAYTAYATHGGNWSNTGYACNTTYGARCAYTTTGTGYGAYTNTNGAYATGYTNATHCCN 360
QY      121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db      361 ACNGNGARCCNTGYCCNGARCCNTNMGNACNTAYGGNTNCCNTGYCAYTGYCCNTTY 420
QY      141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
Db      421 AARGARGGNACNTAYWSNTNCCNAARWSNGARTTYGCGTNCNGAYTNGARYTNCCN 480
QY      161 SerTyrLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysArg 180
Db      481 WSNTGGYTACNACNGGNAAYTAYMGNATHGARWSNGTNTYWSNWSNWSNGNAARMGN 540
QY      181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db      541 YTNGGNTGYATHAARATHGCNCNWSNTYNAARGGNATH 579

```

Search completed: November 18, 2004, 19:48:39
 Job time : 875.071 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 19:19:01 ; Search time 163.836 Seconds
(without alignments)
837.317 Million cell updates/sec

Title: US-10-030-937-9
Perfect score: 1018
Sequence: 1 MSLMQAPLLIALGLLALTP.....LSSGKRLGCIKIAASLKGI 193

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10030937/runat_16112004_153014_2947/app_query.fasta_1.789
-DB=Issued Patents NA -QMT=fastcap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937@cgn_1_200@runat_16112004_153014_2947 -NCPU=6
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97.5	9.6	20966	4 US-09-776-976-7	Sequence 7, Appli
2	97.5	9.6	20966	4 US-09-909-547-7	Sequence 7, Appli
3	97.5	9.6	20966	4 US-09-569-852B-1	Sequence 1, Appli
4	89.5	8.8	890	1 US-08-592-126-78	Sequence 78, Appl
5	89.5	8.8	890	1 US-08-168-595-78	Sequence 78, Appl
6	89.5	8.8	1633	1 US-07-866-979-5	Sequence 5, Appli
7	89.5	8.8	1633	2 US-08-466-906B-5	Sequence 5, Appli
8	89.5	8.8	1633	3 US-08-706-281A-5	Sequence 5, Appli
9	89.5	8.8	1633	3 US-09-201-746-5	Sequence 5, Appli
10	89.5	8.8	1633	3 US-09-097-231-5	Sequence 5, Appli
11	89.5	8.8	1633	4 US-09-353-099-5	Sequence 5, Appli
12	89.5	8.8	1633	4 US-09-016-434-1343	Sequence 1343, Ap

13	88.5	8.7	1388	4 US-09-976-594-806	Sequence 806, App
14	88	8.6	29485	4 US-09-785-381-6	Sequence 6, Appli
15	88	8.6	43950	3 US-09-735-934A-3	Sequence 3, Appli
16	88	8.6	43950	4 US-10-060-332-3	Sequence 3, Appli
17	88	8.6	43950	4 US-10-339-657-3	Sequence 3, Appli
18	86.5	8.5	2781	3 US-09-302-812-7	Sequence 7, Appli
19	86.5	8.5	2781	3 US-09-511-477-7	Sequence 7, Appli
20	86.5	8.5	2781	3 US-09-511-507-7	Sequence 7, Appli
21	86	8.4	588	4 US-09-860-793-2	Sequence 2, Appli
22	86	8.4	29629	4 US-09-729-995-3	Sequence 3, Appli
23	86	8.4	29629	4 US-10-135-689-3	Sequence 3, Appli
24	86	8.4	37030	4 US-08-311-731A-25	Sequence 25, Appli
25	85.5	8.4	1903	6 5466668-5	Patent No. 5466668
26	82.5	8.1	1338	1 US-08-307-444A-8	Sequence 8, Appli
27	82.5	8.1	1338	1 US-08-587-389-8	Sequence 8, Appli
28	82.5	8.1	1368	1 US-08-307-444A-7	Sequence 7, Appli
29	82.5	8.1	1368	1 US-08-587-389-7	Sequence 7, Appli
30	82.5	8.1	1425	1 US-08-307-444A-6	Sequence 6, Appli
31	82.5	8.1	1425	1 US-08-307-444A-23	Sequence 23, Appli
32	82.5	8.1	1425	1 US-08-587-389-6	Sequence 6, Appli
33	82.5	8.1	1425	1 US-08-587-389-23	Sequence 23, Appli
34	82.5	8.1	1428	1 US-08-014-723-3	Sequence 3, Appli
35	82.5	8.1	1428	1 US-08-014-723-4	Sequence 4, Appli
36	82.5	8.1	1428	1 US-08-110-011A-3	Sequence 3, Appli
37	82.5	8.1	1428	1 US-08-110-011A-4	Sequence 4, Appli
38	82.5	8.1	1491	1 US-08-312-870-2	Sequence 2, Appli
39	82.5	8.1	1491	4 US-09-331-793-21	Sequence 21, Appli
40	82.5	8.1	1680	1 US-08-014-723-13	Sequence 13, Appli
41	82.5	8.1	1680	1 US-08-014-723-15	Sequence 15, Appli
42	82.5	8.1	1680	1 US-08-110-011A-13	Sequence 13, Appli
43	82.5	8.1	1680	1 US-08-110-011A-15	Sequence 15, Appli
44	82.5	8.1	2011	1 US-08-170-290A-53	Sequence 53, Appli
45	82.5	8.1	2463	1 US-08-307-444A-10	Sequence 10, Appli

ALIGNMENTS

RESULT 1

US-09-776-976-7
; Sequence 7, Application US/09776976
; Patent No. 6566332
; GENERAL INFORMATION:
; APPLICANT: Fruebis, Joachim
; APPLICANT: Erickson, Mary Ruth
; APPLICANT: Yen, Frances
; TITLE OF INVENTION: ORG3 Globular Head and Uses Thereof for Decreasing Body Mass
; FILE REFERENCE: 76 US4 REG
; CURRENT APPLICATION NUMBER: US/09/776,976
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/758,055
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: US 60/176,228
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/198,087
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: US 60/299,881
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 20966
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..4811
; OTHER INFORMATION: 5' regulatory region
; NAME/KEY: exon
; LOCATION: 4812..4851
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 15144..15365

```

OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 16277..20559
OTHER INFORMATION: exon 3
NAME/KEY: misc feature
LOCATION: 20560..20966
OTHER INFORMATION: 3' regulatory region
NAME/KEY: allele
LOCATION: 3787
OTHER INFORMATION: 9-27-261 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 1118
OTHER INFORMATION: 99-14387-129 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 15120
OTHER INFORMATION: 9-12-48 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 15196
OTHER INFORMATION: 9-12-124 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 15427
OTHER INFORMATION: 9-12-355 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 15500
OTHER INFORMATION: 9-12-428 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 15863
OTHER INFORMATION: 99-14405-105 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 17170
OTHER INFORMATION: 9-16-189 : polymorphic base deletion of A
NAME/KEY: primer_bind
LOCATION: 3528..3545
OTHER INFORMATION: 9-27.pu
NAME/KEY: primer_bind
LOCATION: 3928..3946
OTHER INFORMATION: 9-27.rp complement
NAME/KEY: primer_bind
LOCATION: 10990..11008
OTHER INFORMATION: 99-14387.pu
NAME/KEY: primer_bind
LOCATION: 11423..11442
OTHER INFORMATION: 99-14387.rp complement
NAME/KEY: primer_bind
LOCATION: 15073..15092
OTHER INFORMATION: 9-12.pu
NAME/KEY: primer_bind
LOCATION: 15503..15520
OTHER INFORMATION: 9-12.rp complement
NAME/KEY: primer_bind
LOCATION: 15759..15776
OTHER INFORMATION: 99-14405.pu
NAME/KEY: primer_bind
LOCATION: 16191..16211
OTHER INFORMATION: 99-14405.rp complement
NAME/KEY: primer_bind
LOCATION: 16982..17001
OTHER INFORMATION: 9-16.pu
NAME/KEY: primer_bind
LOCATION: 17384..17402
OTHER INFORMATION: 9-16.rp complement
NAME/KEY: misc_binding
LOCATION: 3775..3799
OTHER INFORMATION: 9-27-261.probe
NAME/KEY: misc_binding
LOCATION: 11106..11130
OTHER INFORMATION: 99-14387-129.probe
NAME/KEY: misc_binding
LOCATION: 15108..15132
OTHER INFORMATION: 9-12-48.probe
NAME/KEY: misc_binding
LOCATION: 15184..15208
OTHER INFORMATION: 9-12-124.probe

```

```

NAME/KEY: misc binding
LOCATION: 15415..15439
OTHER INFORMATION: 9-12-355.probe
NAME/KEY: misc binding
LOCATION: 15488..15512
OTHER INFORMATION: 9-12-428.probe
NAME/KEY: misc binding
LOCATION: 15851..15875
OTHER INFORMATION: 99-14405-105.probe
NAME/KEY: misc binding
LOCATION: 17158..17182
OTHER INFORMATION: 9-16-189.probe
NAME/KEY: primer_bind
LOCATION: 3768..3786
OTHER INFORMATION: 9-27-261.mis
NAME/KEY: primer_bind
LOCATION: 3788..3806
OTHER INFORMATION: 9-27-261.mis complement
NAME/KEY: primer_bind
LOCATION: 11099..11117
OTHER INFORMATION: 99-14387-129.mis
NAME/KEY: primer_bind
LOCATION: 11119..11137
OTHER INFORMATION: 99-14387-129.mis complement
NAME/KEY: primer_bind
LOCATION: 15101..15119
OTHER INFORMATION: 9-12-48.mis
NAME/KEY: primer_bind
LOCATION: 15121..15139
OTHER INFORMATION: 9-12-48.mis complement
NAME/KEY: primer_bind
LOCATION: 15177..15195
OTHER INFORMATION: 9-12-124.mis
NAME/KEY: primer_bind
LOCATION: 15197..15215
OTHER INFORMATION: 9-12-124.mis complement
NAME/KEY: primer_bind
LOCATION: 15408..15426
OTHER INFORMATION: 9-12-355.mis
NAME/KEY: primer_bind
LOCATION: 15428..15446
OTHER INFORMATION: 9-12-355.mis complement
NAME/KEY: primer_bind
LOCATION: 15481..15499
OTHER INFORMATION: 9-12-428.mis
NAME/KEY: primer_bind
LOCATION: 15501..15519
OTHER INFORMATION: 9-12-428.mis complement
NAME/KEY: primer_bind
LOCATION: 15844..15862
OTHER INFORMATION: 99-14405-105.mis
NAME/KEY: primer_bind
LOCATION: 15864..15882
OTHER INFORMATION: 99-14405-105.mis complement
NAME/KEY: primer_bind
LOCATION: 17151..17169
OTHER INFORMATION: 9-16-189.mis
NAME/KEY: primer_bind
LOCATION: 17171..17189
OTHER INFORMATION: 9-16-189.mis complement
US-09-776-976-7
Alignment Scores:
Pred. No.: 5.31 Length: 20966
Score: 97.50 Matches: 57
Percent Similarity: 41.75% Conservative: 24
Best Local Similarity: 29.38% Mismatches: 72
Query Match: 9.58% Indels: 41
DB: 4 Gaps: 10
US-10-030-937-9 (1-193) x US-09-776-976-7 (1-20966)
Qy 19 ThrProAlaGlnAla-----HisLeuLysLysProSerGlnLeuSerSer 33

```

```
Db 6579 ACTCAGCTGGGCAAAAGACCAAACTCCATCTCAAAAAAATAATGACACA 6638
Qy 34 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArgSerLeuThrLeu 53
Db 6639 AGACTGGCTCTGTCTTTTGGGGA-----CAGGGTCTCACTCTA 6680
Qy 54 GluProAspProIleValValProGlyAsnValThr-----LeuSerValVal 69
Db 6681 TCACCC---AGCTCGAGTGCAGTGGTCAATCACAGCTCACTGCAGGCTCGAATTCACA 6737
Qy 70 GlySerThrSerValProLeuSerSer-ProLeuLysValAspLeuValLeuGlySgl 89
Db 6738 GGCTCAAGTACCCCTCCATCTTACCTCCTGAGTAGCTGGGACTACAGGTGTGGCAAC 6797
Qy 89 uValAlaGlyLeuThrIleLysIleProCysThrAspTyrIleGlySerCysThrPheGl 109
Db 6798 CATGCTGGCTAAATTTTAAAAAATTTTGTAGAGATGAGGTCTCACTA----- 6846
Qy 109 uHisPheCysAspValLeuAspMetLeuLeuProThrGlyGluProCysProGluProLe 129
Db 6847 -TATTGGCTGGGGGCGCTCAAACTCCTGGGCTCAGCAGTCCCTCCA---CCTCAGCCCTCC 6902
Qy 129 uArgThrTyrGlyLeu---ProCysHis-----CysProPheLysGluGlyTh 144
Db 6903 CAAGAAGCTGGGATATATGCTTCTTTTAAGTGGCTGTAGGACAACTTTCAC 6962
Qy 144 rTyrSerLeuProLys-SerGluPheAlaValProAspLeuGluLeuProSer----- 161
Db 6963 CTACTCTTGTCAAGCCAGTGGACCGGTGGTCCACAGACATACGGCTAAAGTCAAGAGGTG 7022
Qy 162 -----TrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyL 179
Db 7023 ATGCTTTTGGAGAGATCTTCAATCAGGAATTTCAATCAGAAATTCATCATGTGGAG 7082
Qy 179 ysArgLeu-----GlyCys 183
Db 7083 AGAGACTATCTTAAATATGGTGGTGGTGGGATGC 7120

RESULT 2
US-09-909-547-7
; Sequence 7, Application US/09909547
; Patent No. 6579852
; GENERAL INFORMATION:
; APPLICANT: Fruebis, Joachim
; APPLICANT: Erickson, Mary Ruth
; APPLICANT: Yen, Frances
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: OBG3 Globular Head and Uses Thereof for Decreasing Body Mass
; FILE REFERENCE: 76 US6 CIP
; CURRENT APPLICATION NUMBER: US/09/909,547
; CURRENT FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 09/776,976
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/758,055
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: US 60/299,881
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 60/198,087
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: US 60/176,228
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent.Pm
; SEQ ID NO 7
; LENGTH: 20966
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..4811
; OTHER INFORMATION: 5' regulatory region
; NAME/KEY: exon
; LOCATION: 4812..4851
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 15144..15365
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 16277..20559
; OTHER INFORMATION: exon 3
; NAME/KEY: misc_feature
; LOCATION: 20560..20966
; OTHER INFORMATION: 3' regulatory region
; NAME/KEY: allele
; LOCATION: 3787
; OTHER INFORMATION: 9-27-261 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 11118
; OTHER INFORMATION: 99-14387-129 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 15120
; OTHER INFORMATION: 9-12-48 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 15196
; OTHER INFORMATION: 9-12-124 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 15427
; OTHER INFORMATION: 9-12-355 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 15500
; OTHER INFORMATION: 9-12-428 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 15863
; OTHER INFORMATION: 99-14405-105 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 17170
; OTHER INFORMATION: 9-16-189 : polymorphic base deletion of A
; NAME/KEY: primer_bind
; LOCATION: 3528..3545
; OTHER INFORMATION: 9-27.pu
; NAME/KEY: primer_bind
; LOCATION: 3928..3946
; OTHER INFORMATION: 9-27.rp complement
; NAME/KEY: primer_bind
; LOCATION: 10990..11008
; OTHER INFORMATION: 99-14387.pu
; NAME/KEY: primer_bind
; LOCATION: 11423..11442
; OTHER INFORMATION: 99-14387.rp complement
; NAME/KEY: primer_bind
; LOCATION: 15073..15092
; OTHER INFORMATION: 9-12.pu
; NAME/KEY: primer_bind
; LOCATION: 15503..15520
; OTHER INFORMATION: 9-12.rp complement
; NAME/KEY: primer_bind
; LOCATION: 15759..15776
; OTHER INFORMATION: 99-14405.pu
; NAME/KEY: primer_bind
; LOCATION: 16191..16211
; OTHER INFORMATION: 99-14405.rp complement
; NAME/KEY: primer_bind
; LOCATION: 16982..17001
; OTHER INFORMATION: 9-16.pu
; NAME/KEY: primer_bind
; LOCATION: 17384..17402
; OTHER INFORMATION: 9-16.rp complement
; NAME/KEY: misc_binding
; LOCATION: 3775..3799
; OTHER INFORMATION: 9-27-261.probe
; NAME/KEY: misc_binding
; LOCATION: 11106..11130
; OTHER INFORMATION: 99-14387-129.probe
; NAME/KEY: misc_binding
; LOCATION: 15108..15132
```

```

OTHER INFORMATION: 9-12-48.probe
NAME/KEY: misc binding
LOCATION: 15184..15208
OTHER INFORMATION: 9-12-124.probe
NAME/KEY: misc binding
LOCATION: 15415..15439
OTHER INFORMATION: 9-12-355.probe
NAME/KEY: misc binding
LOCATION: 15488..15512
OTHER INFORMATION: 9-12-428.probe
NAME/KEY: misc binding
LOCATION: 15851..15875
OTHER INFORMATION: 99-14405-105.probe
NAME/KEY: misc binding
LOCATION: 17158..17182
OTHER INFORMATION: 9-16-189.probe
NAME/KEY: primer bind
LOCATION: 3768..3786
OTHER INFORMATION: 9-27-261.mis
NAME/KEY: primer bind
LOCATION: 3788..3806
OTHER INFORMATION: 9-27-261.mis complement
NAME/KEY: primer bind
LOCATION: 11099..11117
OTHER INFORMATION: 99-14387-129.mis
NAME/KEY: primer bind
LOCATION: 11119..11137
OTHER INFORMATION: 99-14387-129.mis complement
NAME/KEY: primer bind
LOCATION: 15101..15119
OTHER INFORMATION: 9-12-48.mis
NAME/KEY: primer bind
LOCATION: 15121..15139
OTHER INFORMATION: 9-12-48.mis complement
NAME/KEY: primer bind
LOCATION: 15177..15195
OTHER INFORMATION: 9-12-124.mis
NAME/KEY: primer bind
LOCATION: 15197..15215
OTHER INFORMATION: 9-12-124.mis complement
NAME/KEY: primer bind
LOCATION: 15408..15426
OTHER INFORMATION: 9-12-355.mis
NAME/KEY: primer bind
LOCATION: 15481..15499
OTHER INFORMATION: 9-12-428.mis
NAME/KEY: primer bind
LOCATION: 15501..15519
OTHER INFORMATION: 9-12-428.mis complement
NAME/KEY: primer bind
LOCATION: 15844..15862
OTHER INFORMATION: 99-14405-105.mis
NAME/KEY: primer bind
LOCATION: 15864..15882
OTHER INFORMATION: 99-14405-105.mis complement
NAME/KEY: primer bind
LOCATION: 17151..17169
OTHER INFORMATION: 9-16-189.mis
NAME/KEY: primer bind
LOCATION: 17171..17189
OTHER INFORMATION: 9-16-189.mis complement
US-09-909-547-7
Alignment Scores:
Pred. No.: 5.31 Length: 20966
Score: 97.50 Matches: 57
Percent Similarity: 41.75% Conservative: 24
Best Local Similarity: 29.38% Mismatches: 72
Query Match: 9.58% Indels: 41
DB: 4 Gaps: 10

```

```

US-10-030-937-9 (1-193) x US-09-909-547-7 (1-20966)
QY 19 ThrProAlaGlnAla-----HisLeuLysLysProSerGlnLeuSerSer 33
|||||
Db 6579 ACTCCAGCCTGGGCAAAAGAGCAAACTCCATCTCAAAAAAATAGACACA 6638
|||||
QY 34 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArgSerLeuThrLeu 53
|||||
Db 6639 AGACTGGCTCCTTTGCTTTTGGGA-----CAGGGCTCACTCTA 6680
|||||
QY 54 GluProAspProIleValProGlyAsnValThr-----LeuSerValVal 69
|||||
Db 6681 TCACCC---AGCTGGAGTCAGTGTGCAATCACAGCTCAGTGCAGCTCGATTTCCCA 6737
|||||
QY 70 GlySerThrSerValProLeuSerSer-ProLeuLysValAspLeuValLeuGluLysG1 89
|||||
Db 6738 GGCTCAAGTGACCTCCCATCTTAGCTCTCTAGTAGCTGGACTACAGGTGTGTGCAAC 6797
|||||
QY 89 uValAlaGlyLeuTrpIleLysIleProCysThrAspTyrIleGlySerCysThrPheG1 109
|||||
Db 6798 CATGCCCTGGCTAAATTTTAAATAATTTTGTAGATGAGTCTCACTA----- 6846
|||||
QY 109 uHisPheCysAspValLeuAspMetLeuIleProThrGlyGluProCysProGluProLe 129
|||||
Db 6847 -TAATGGCTGGGGGCTCAAACTCTGGCTCAGCAGTCTCCCA---CCTCAGCTCC 6902
|||||
QY 129 uArgThrTyrGlyLeu---ProCysHis-----CysProPheLysGluGlyTh 144
|||||
Db 6903 CAAAAGCTGGGATTATATGCTTCTTTTAAAGTGGCTGTAGGACAAACTTTCCAC 6962
|||||
QY 144 rTyrSerLeuProLys-SerGluPheAlaValProAspLeuGluLeuProSer----- 161
|||||
Db 6963 CTACTCCTTGTCAAGCAGTGGACCGGTGTCCAGACATACGGCTAAAGTCAAGAGGTG 7022
|||||
QY 162 -----TrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerglyL 179
|||||
Db 7023 ATGCTCTTTGGAGAGATACTTTCAATCAGGAATTTCAATCAGAAATCAATCATGTGGAG 7082
|||||
QY 179 ysArgLeu-----GlyCys 183
|||||
Db 7083 AGAGACTTATCTATAAAATGTGTGTGTGGTGGATGC 7120
|||||
RESULT 3
US-09-569-852B-1
; Sequence 1, Application US/09569852B
; Patent No. 6582909
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Denison, Blake
; APPLICANT: Yen-Potin, Frances
; TITLE OF INVENTION: APM1 Biallelic Markers and Uses Thereof
; FILE REFERENCE: GEN-T113XC2
; CURRENT APPLICATION NUMBER: US/09/569,852B
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: PCT/IB99/01858
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: US 09/434,848
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: US 60/119,593
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/107,113
; PRIOR FILING DATE: 1998-11-04
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 20966
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(4811)

```

```
OTHER INFORMATION: 5' regulatory region
NAME/KEY: primer_bind
LOCATION: (14683)..(14701)
OTHER INFORMATION: 17-34-860.mis
NAME/KEY: primer_bind
LOCATION: (14703)..(14721)
OTHER INFORMATION: 17-34-860.mis complement
NAME/KEY: primer_bind
LOCATION: (14738)..(14756)
OTHER INFORMATION: 17-34-915.mis
NAME/KEY: primer_bind
LOCATION: (14758)..(14776)
OTHER INFORMATION: 17-34-915.mis complement
NAME/KEY: primer_bind
LOCATION: (14796)..(14814)
OTHER INFORMATION: 17-35-71.mis
NAME/KEY: primer_bind
LOCATION: (14816)..(14834)
OTHER INFORMATION: 17-35-71.mis complement
NAME/KEY: primer_bind
LOCATION: (15031)..(15049)
OTHER INFORMATION: 17-35-306.mis
NAME/KEY: primer_bind
LOCATION: (15051)..(15069)
OTHER INFORMATION: 17-35-306.mis complement
NAME/KEY: primer_bind
LOCATION: (15101)..(15119)
OTHER INFORMATION: 9-12-48.mis
NAME/KEY: primer_bind
LOCATION: (15121)..(15139)
OTHER INFORMATION: 9-12-48.mis complement
NAME/KEY: primer_bind
LOCATION: (15177)..(15195)
OTHER INFORMATION: 9-12-124.mis
NAME/KEY: primer_bind
LOCATION: (15197)..(15215)
OTHER INFORMATION: 9-12-124.mis complement
NAME/KEY: primer_bind
LOCATION: (15408)..(15426)
OTHER INFORMATION: 9-12-355.mis
NAME/KEY: primer_bind
LOCATION: (15428)..(15446)
OTHER INFORMATION: 9-12-355.mis complement
NAME/KEY: primer_bind
LOCATION: (15481)..(15499)
OTHER INFORMATION: 9-12-428.mis
NAME/KEY: primer_bind
LOCATION: (15501)..(15519)
OTHER INFORMATION: 9-12-428.mis complement
NAME/KEY: primer_bind
LOCATION: (15661)..(15679)
OTHER INFORMATION: 17-36-47.mis
NAME/KEY: primer_bind
LOCATION: (15681)..(15699)
OTHER INFORMATION: 17-36-47.mis complement
NAME/KEY: primer_bind
LOCATION: (15771)..(15789)
OTHER INFORMATION: 17-36-120.mis
NAME/KEY: primer_bind
LOCATION: (15791)..(15809)
OTHER INFORMATION: 17-36-120.mis complement
NAME/KEY: primer_bind
LOCATION: (15844)..(15862)
OTHER INFORMATION: 99-14405-105.mis
NAME/KEY: primer_bind
LOCATION: (15864)..(15882)
OTHER INFORMATION: 99-14405-105.mis complement
NAME/KEY: primer_bind
LOCATION: (17151)..(17169)
OTHER INFORMATION: 9-16-189.mis
NAME/KEY: primer_bind
LOCATION: (17171)..(17189)
OTHER INFORMATION: 9-16-189.mis complement

NAME/KEY: primer_bind
LOCATION: (17810)..(17828)
OTHER INFORMATION: 17-37-629.mis
NAME/KEY: primer_bind
LOCATION: (17830)..(17848)
OTHER INFORMATION: 17-37-629.mis complement
NAME/KEY: primer_bind
LOCATION: (17992)..(18010)
OTHER INFORMATION: 17-37-811.mis
NAME/KEY: primer_bind
LOCATION: (18012)..(18030)
OTHER INFORMATION: 17-37-811.mis complement
NAME/KEY: primer_bind
LOCATION: (18470)..(18488)
OTHER INFORMATION: 17-38-349.mis
NAME/KEY: primer_bind
LOCATION: (18490)..(18508)
OTHER INFORMATION: 17-38-349.mis complement
NAME/KEY: primer_bind
LOCATION: (926)..(944)
OTHER INFORMATION: 17-30-216.mis
NAME/KEY: primer_bind
LOCATION: (946)..(964)
OTHER INFORMATION: 17-30-216.mis complement
NAME/KEY: primer_bind
LOCATION: (3719)..(3737)
OTHER INFORMATION: 9-27-211.mis
NAME/KEY: primer_bind
LOCATION: (3739)..(3757)
OTHER INFORMATION: 9-27-211.mis complement
NAME/KEY: primer_bind
LOCATION: (3754)..(3772)
OTHER INFORMATION: 9-27-246.mis
NAME/KEY: primer_bind
LOCATION: (3774)..(3792)
OTHER INFORMATION: 9-27-246.mis complement
NAME/KEY: primer_bind
LOCATION: (3768)..(3786)
OTHER INFORMATION: 9-27-261.mis
NAME/KEY: primer_bind
LOCATION: (3788)..(3806)
OTHER INFORMATION: 9-27-261.mis complement
NAME/KEY: primer_bind
LOCATION: (5076)..(5094)
OTHER INFORMATION: 17-31-298.mis
NAME/KEY: primer_bind
LOCATION: (5096)..(5114)
OTHER INFORMATION: 17-31-298.mis complement
NAME/KEY: primer_bind
LOCATION: (5191)..(5209)
OTHER INFORMATION: 17-31-413.mis
NAME/KEY: primer_bind
LOCATION: (5211)..(5229)
OTHER INFORMATION: 17-31-413.mis complement
NAME/KEY: primer_bind
LOCATION: (5364)..(5382)
OTHER INFORMATION: 17-31-413.mis complement
NAME/KEY: primer_bind
LOCATION: (10618)..(10636)
OTHER INFORMATION: 17-32-24.mis
NAME/KEY: primer_bind
LOCATION: (10638)..(10656)
OTHER INFORMATION: 17-32-24.mis complement
NAME/KEY: primer_bind
LOCATION: (11020)..(11038)
OTHER INFORMATION: 99-14387-50.mis
NAME/KEY: primer_bind
LOCATION: (11040)..(11058)
OTHER INFORMATION: 99-14387-50.mis complement
NAME/KEY: primer_bind
LOCATION: (11099)..(11117)
OTHER INFORMATION: 99-14387-129.mis
NAME/KEY: primer_bind
```

```

; LOCATION: (11119)..(11137)
; OTHER INFORMATION: 99-14387-129.mis complement
; NAME/KEY: primer bind
; LOCATION: (11169)..(11187)
; OTHER INFORMATION: 99-14387-199.mis
; NAME/KEY: primer bind
; LOCATION: (11189)..(11207)
; OTHER INFORMATION: 99-14387-199.mis complement
; NAME/KEY: primer bind
; LOCATION: (13954)..(13972)
; OTHER INFORMATION: 17-33-TGAGACT.mis
; NAME/KEY: primer bind
; LOCATION: (13974)..(13992)
; OTHER INFORMATION: 17-33-TGAGACT.mis complement
; NAME/KEY: exon
; LOCATION: (4812)..(4851)
; OTHER INFORMATION:
; NAME/KEY: exon
; LOCATION: (15144)..(15365)
; OTHER INFORMATION:
; NAME/KEY: exon
; LOCATION: (16277)..(20559)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (20560)..(20966)
; OTHER INFORMATION: 3' regulatory region

```

```

Alignment Scores:
Pred. No.: 5.31 Length: 20966
Score: 97.50 Matches: 57
Percent Similarity: 41.75% Conservative: 24
Best Local Similarity: 29.38% Mismatches: 72
Query Match: 9.58% Indels: 41
DB: 4 Gaps: 10

```

US-10-030-937-9 (1-193) x US-09-569-852B-1 (1-20966)

```

QY 19 ThrProAlaGlnla-----HisLeuLysProSerGlnLeuSerSer 33
DB 6579 ACTCCAGCTGGCAAAAGAGCAAACTCCATCTCAAAAAAATATAGACACA
QY 34 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArgSerLeuThrLeu 53
DB 6639 AGACTGCTCTCTCTTTTGGGA-----CAGGGTCTCACTCTA 6680
QY 54 GluProAspProIleValValProGlyAsnValThr-----LeuSerValVal 69
DB 6681 TCACCC---AGGCTGGAGTGCAGTGTGCAATCACAGCTCACTCGACCTCGATTCCCA 6737
QY 70 GlySerThrSerValProLeuSerSer-ProLeuLysValAspLeuValLeuGluLysG1 89
DB 6738 GGCTCAAGTGACCTCCATCTTAGCTCTGAGTAGCTGGGACTCAGGTGTGTGCAAC 6797
QY 89 uAlaAlaGlyLeuTrpIleLysIleProCysThrAspTyrIleGlySerCysThrPheG1 109
DB 6798 CATGCCCTGGCTAATTTTAAATTTTGTAGAGATGAGTCTCACTA----- 6846
QY 109 uHisPheCysAspValLeuAspMetLeuIleProThrGlyGluProCysProGluProLe 129
DB 6847 -TATTGCTGGGGGCGCTCAAACTCTCGGCTCAGCAGTCTCCCA---CCTCAGCCTCC 6902
QY 129 uArgThrTyrClyLeu---ProCysHis-----CysProPheLysGluGlyTh 144
DB 6903 CAAAGAGCTGGGATATATGCTTCTCTTTTAAAGTGGCTGTAGGACAAACTTTCCAC 6962
QY 144 rTyrSerLeuProLys-SerGluPheAlaValProAspLeuGluLeuProSer----- 161
DB 6963 CTACTCTTTTCAAGCCAGTGGACCGGTGGTCCAGACATACGCTAAAGTCAAGAGGTG 7022
QY 162 -----TrpLeuThrThrClyAsnTyrArgIleGluSerValLeuSerSerSerglyL 179
DB 7023 ATGCTCTTTTGAGAGATACTTTCAATCAGGAATTTCAATCAGAAATTTCAATCATGTGGAG 7082

```

```

QY 179 ysArgLeu-----GlyCys 183
DB 7083 AGAGACTTATCTATAAAATGTGCTGCTGGTGGATGC 7120

RESULT 4
US-08-592-126-78
; Sequence 78, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 890 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: G256.seq
; US-08-592-126-78

```

```

Alignment Scores:
Pred. No.: 0.274 Length: 890
Score: 89.50 Matches: 53
Percent Similarity: 39.89% Conservative: 22
Best Local Similarity: 28.19% Mismatches: 67
Query Match: 8.79% Indels: 47
DB: 1 Gaps: 10

```

US-10-030-937-9 (1-193) x US-08-592-126-78 (1-890)

```

QY 3 SerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuAlaThrProAlaGln 22
DB 49 TCCCTTTACCAAGAGCCCTTA-----CCCATGGGGTGGGTACAGGAGGCCCAAGACAG 102
QY 23 AlaHisLeuLysLysProSerGlnLeuSerSerPheSer---TrpAspAsnCysPheGlu 41
DB 103 GCCCGTATCAGGAGACCCCTCTTCTCAGGGGCTGCCCTCTGGGATAC----- 153
QY 42 GlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValValPro 61
DB 154 -----CACCCCGCCCTTCTGGGTTTCCT 177
QY 62 GlyAsnValThrLeuSerValValGlySerThrSerVal----- 74

```


178	GCTTCCTAT---	CTGGCTGACAGTTTCTCAGTGCCCTCTGTGGATTTCCCATGGTCTGTCC	234
QY			
75	ProLeuSerSerProLeuLysValAspLeuValLeuGluLysGluValAlaGlyLeuTrp	94	
235	CCACTCACATCCCTCTCTGCAAACTTGCTACTG-----GGCTGTGCAC	279	
Db			
95	IleLysIleProCysThrAspTyrIleGlySerCysThrPheGluHisPheCysAspVal	114	
280	CTGGCAAAATCCATGCTCAGCACAGAGGGGATCAAGACCTCTCAATACAACATGT-----	333	
Db			
115	LeuAspMetIleuIleProThrGlyGluProCysProGluProLeuArgThrTyrGlyLeu	134	
334	-----CTCCCTGGCCAAT-----CCCTGGCCCCAGCAGCTGAGGCCCATCTGAAA	377	
Db			
135	ProCysHisCys---ProPheLysGlu-GlyThrTyrSerLeuProLysSerGluPheAl	153	
QY			
378	CCAGGGAGTTGCTCTCTTTCTCCTCCCTTGACCTCACCCCTCAGACCCATGCAATTCTG	437	
Db			
153	aValProAspLeu-----GluLeuProSerTrpLeuThrGlyAs	167	
QY			
438	CCTCTAAACCTCCCGAGCCAGCCCTCTCCCGAGCTCCAGTGACAGTGCTCTCAGGTAC	497	
Db			
167	nTyrArgIleGluSerValLeu	174	
QY			
498	CTGAGCTCAGCTCTCGGTGCTTA	519	
Db			

RESULT 5

```

US-09-168-595-78
; Sequence 78, Application US/09168595
; Patent No. 6555666
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/168,595
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,126
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 890 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: G256.seq
; US-09-168-595-78

```

Alignment Scores:					
Pred. No.:	0.274	Length:	890		
Score:	89.50	Matches:	53		
Percent Similarity:	39.89%	Conservative:	22		
Best Local Similarity:	28.19%	Mismatches:	67		
Query Match:	8.79%	Indels:	47		
DB:	4	Gaps:	10		
 US-10-030-937-9 (1-193) x US-09-168-595-78 (1-890)					
QY	3	SerLeuMetGlnAlaProLeuLysLeuAlaLeuGlyLeuLeuLeuAlaThrProAlaGln	22		
Db	49	TCCCTTTACCAAGGCCCTA-----CCCATGGGTGGTGTCAGCAGCGCCCAACAGAC	102		
QY	23	AlaHisLeuLysLysProSerGlnLeuSerSerPheSer---TrpAspAsnCysPheGlu	41		
Db	103	GCCCGTATCAGGAGCACCCCTCTCTCTCAGGGGCTGCCCTCTGGATAAC-----	153		
QY	42	GlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValValPro	61		
Db	154	-----CACCCTCCGCCTCTCTGGGTTCCT	177		
QY	62	GlyAsnValThrLeuSerValValGlySerThrSerVal-----	74		
Db	178	GCTTTCAT--CTGGCTGCAGTTCTCAGGTCCTCTTGGAATTTCCCATGGTCTGTCC	234		
QY	75	ProLeuSerSerProLeuLysValAspLeuValLeuGluLysGluValAlaGlyLeuTrp	94		
Db	235	CCACTACATCCCTCTCTGCAAACCTTGCTACTG-----GGCTGTGCAC	279		
QY	95	IleLysIleProCysThrAspTyrlIleGlySerCysThrPheGluHisPheCysAspVal	114		
Db	280	CTGGCAAAATCCATGTCTCAGCACAGCGGGATCAAGACCTCTCAATACAACCTGT---	333		
QY	115	LeuaspMetLeuIleProThrGlyLysProGluProCysProGluProLeuArgThrTyrglyLeu	134		
Db	334	-----CTCCTGCCAAAT-----CCCTGCCCCAGACGCTCAGGCCCCAGTCTGAAA	377		
QY	135	ProCysHisCys--ProPheLysGlu-GlyThrTyr-SerLeuProLysSerGluPheAl	153		
Db	378	CAGGAGATTGTCTCTCTTCTCCTCTGACCTCACCCCTCAGACCAATGCCAATCTTG	437		
QY	153	ValProAspLeu-----GluLeuProSerTrpLeuThrThrGlyAs	167		
Db	438	CCTCCTAAACCTCCAGCGCCAGCCCTCCCCCAGCTCCAGTGACAGTGTCCTCAGGTAC	497		
QY	167	nTyrArgIleGluSerValLeu	174		
Db	498	CTGAGCTCAGCTCTCGGTGCTA	519		

RESULT 6

```

US-07-866-979/c
Sequence 5, Application US/07866979
Patent No. 5532347
GENERAL INFORMATION:
APPLICANT: Cone, Roger D
APPLICANT: Mountjoy, Kathleen G
TITLE OF INVENTION: Melanocyte Stimulating Hormone Receptor
TITLE OF INVENTION: and Uses
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0.

```

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/866,979
 ; FILING DATE: 19920410
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: No. 5532347nan, Kevin E
 ; REGISTRATION NUMBER: 35,303
 ; REFERENCE/DOCKET NUMBER: 92,154
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312-715-1000
 ; TELEFAX: 312-715-1234
 ; TELEX: 910-221-5317
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1633 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 462..1415
 ; FEATURE:
 ; NAME/KEY: 5'UTR
 ; LOCATION: 1..461
 ; FEATURE:
 ; NAME/KEY: 3'UTR
 ; LOCATION: 1416..1633
 ; US-07-866-979-5

Alignment Scores:
 Pred. No.: 0.752 Length: 1633
 Score: 89.50 Matches: 49
 Percent Similarity: 38.25% Conservative: 21
 Best Local Similarity: 26.78% Mismatches: 62
 Query Match: 8.79% Indels: 51
 DB: 1 Gaps: 8

US-10-030-937-9 (1-193) x US-07-866-979-5 (1-1633)

QY 26 LysLysProSerGlnLeu-----SerSer 33
 DB 595 AAGAGCCGTCAGAGATGGACCTCCAGGACCGGGCTCTGCTGTTGGCAGCCAGC 536
 QY 34 PheSerTrpAspLeu-----PheGluGlyLysAspProAlaValIleArgSerLeu 51
 DB 535 CCCAGCTGGGGGATGGCTGTGGGGGTGGAGTTCAGGGAGGCC-----AGAAAGTCTT 485
 QY 52 ThrLeuGluProAspProIleValProGlyAsnValThrLeuSerValValGlySer 71
 DB 484 CTCGGGATCCCTGCACAGCATAGTCTGTCCAGGAGGAGGAGGAGTGTGGAGC 425
 QY 72 ThrSerValProLeuSerSerProLeuValAspLeuValLeuGluLysGluValAla 91
 DB 424 CTCAGGATCCCAAGTCTCTCCCTCCAGGTGCTGTAGTTCATGTGTGCTGCCAGG 365
 QY 92 GlyLeuTrpIleProCysThrAspTyrIleGlySerCysThrPheGluHisPhe 111
 DB 364 GGGCCTTGGGTGCC-----CCATGC----- 344
 QY 112 CysAspValLeuAspMet-LeuIleProThrGlyGluPro-----CysPro-GluPro 128
 DB 343 -----CTGGCTCTCTTCATCTGGGACCCCGACGATCTGCCCTCACACCT 299
 QY 129 LeuArgThrTyr---GlyLeuProCysHisCysProPheLysGluGlyThrTyr----- 145
 DB 298 CTCCTGGCTGGACAGGTTCAGCCCGAGGATGGCCCGACGATCCAGACACCTCTGGCAT 239
 QY 146 -----SerLeuProLysSerGluPheAlaValProAspLeuGlu 158
 DB 238 CAACCGCTGGGCTCAGGATTCACACCAAGCCGTGGCTCAGTCCCTCCCTCCCT 179
 QY 159 LeuProSerTrpLeu-ThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSergl 178

Db 178 CTTCCTGGTGGCTGCTCAGTTCCAGGACATTCACACCTTGGACGCTCTGG 119
 QY 178 Y 178
 Db 118 C 118
 RESULT 7
 US-08-466-906B-5/c
 ; Sequence 5, Application US/08466906B
 ; Patent No. 5849871
 ; GENERAL INFORMATION:
 ; APPLICANT: Cone, Roger D
 ; APPLICANT: Mountjoy, Kathleen G
 ; TITLE OF INVENTION: Melanocyte Stimulating Hormone Receptor
 ; TITLE OF INVENTION: and Uses
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
 ; STREET: 300 South Wacker Drive
 ; CITY: Chicago
 ; STATE: IL
 ; COUNTRY: USA
 ; ZIP: 60606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/466,906B
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: No. 5849871nan, Kevin E
 ; REGISTRATION NUMBER: 35,303
 ; REFERENCE/DOCKET NUMBER: 92,154-H
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312-913-0001
 ; TELEFAX: 312-913-0002
 ; TELEX:

INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1633 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 FEATURE:
 NAME/KEY: 5'UTR
 LOCATION: 1..461
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 462..1415
 FEATURE:
 NAME/KEY: 3'UTR
 LOCATION: 1416..1633
 US-08-466-906B-5
 Alignment Scores:
 Pred. No.: 0.752 Length: 1633
 Score: 89.50 Matches: 49
 Percent Similarity: 38.25% Conservative: 21
 Best Local Similarity: 26.78% Mismatches: 62
 Query Match: 8.79% Indels: 51
 DB: 2 Gaps: 8

US-10-030-937-9 (1-193) x US-08-466-906B-5 (1-1633)

QY 26 LysLysProSerGlnLeu-----SerSer 33
 DB 595 AAGAGCCGTCAGAGATGGACCTCCAGGACCGGGCTCTGCTGTTGGCAGCCAGC 536

QY 34 PheSerTrpAspAsnCys-----PheGluGlyLysAspProAlaValIleArgSerLeu 51
Db 535 CCCAGCTGGGGATGGCTGTGGGGTGGAGTTGAGGGAGCCC-----AGAAAGTCTT 485
QY 52 ThrLeuGluProAspProIleValValProGlyAsnValThrLeuSerValValGlySer 71
Db 484 CTCCTGGATCCCTGCACAGCCATAGTCTCTGCCAGGAGCAGGAGGAGTCTGTGGAGGC 425
QY 72 ThrSerValProLeuSerSerProLeuLysValAspLeuValLeuGluLysGluValAla 91
Db 424 CTCACAGGTCCCAAGTCTTCCCTCCAGGTGTCTCTAGTTCATGCTGCTGCCAGG 365
QY 92 GlyLeuTrpIleLysLeuProCysThrAspTyrIleGlySerCysThrPheGluHisPhe 111
Db 364 GGGCTTGGGTGCC-----CCATGC----- 344
QY 112 CysAspValLeuAspMet-LeuIleProThrGlyGluPro-----CysPro--GluPro 128
Db 343 -----CTGCCCTCTCCATCTGGGCACCCAGGATCTGCCCTCACACCT 299
QY 129 LeuArgThrTyr--GlyLeuProCysHisCysProPheLysGluGlyThrTyr----- 145
Db 298 CTCCTGGCTGGACAGGTCTTCCCTCCAGGTGTCTCTAGTTCATGCTGCTGCCAGG 365
QY 146 -----SerLeuProLysSerGluPheAlaValProAspLeuGlu 158
Db 238 CAACCGCTGGGCTCAGGGATTCACAAACAGCTGGTCTCAGCTCCCTCCCTGCT 179
QY 159 LeuProSerTrpLeu-ThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlu 178
Db 178 CTCCTGGTGGCTGCTGCTCAGGTTCACAGGATTCAGGATTCAGGACCTCTGGACCTCTGG 119
QY 178 Y 178
Db 118 C 118

RESULT 8

US-08-706-281A-5/c
; Sequence 5, Application US/08706281A
; Patent No. 6100048
; GENERAL INFORMATION:
; APPLICANT: Cone, Roger D
; APPLICANT: Fan, Wei
; APPLICANT: Boston, Bruce A
; APPLICANT: Resterton, Robert A
; APPLICANT: Lu, Dongsi
; APPLICANT: Chen, Wenbiao
; TITLE OF INVENTION: Methods and Reagents for Discovering and
; ADDRESS: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/706.281A
; FILING DATE: 04-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6100048nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 96,886
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001

; TELEFAX: 312-913-0002
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1633 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..461
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 462..1415
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 1416..1633
US-08-706-281A-5
Alignment Scores:
Pred. No.: 0.752 Length: 1633
Score: 89.50 Matches: 49
Percent Similarity: 38.25% Conservative: 21
Best Local Similarity: 26.78% Mismatches: 62
Query Match: 8.79% Indels: 51
DB: 3 Gaps: 8

US-10-030-937-9 (1-193) x US-08-706-281A-5 (1-1633)

QY 26 LysLysProSerGlnLeu-----SerSer 33
Db 595 AAGAGCCGTCAGAGATGACACCTCCAGGACCCGGGCTCTGTCTGTGGTGGAGCCAGC 536
QY 34 PheSerTrpAspAsnCys-----PheGluGlyLysAspProAlaValIleArgSerLeu 51
Db 535 CCCAGCTGGGGATGGCTGTGGGGTGGAGTTGAGGGAGCCC-----AGAAAGTCTT 485
QY 52 ThrLeuGluProAspProIleValValProGlyAsnValThrLeuSerValValGlySer 71
Db 484 CTCCTGGATCCCTGCACAGCCATAGTCTCTGCCAGGAGCAGGAGGAGTCTGTGGAGGC 425
QY 72 ThrSerValProLeuSerSerProLeuLysValAspLeuValLeuGluLysGluValAla 91
Db 424 CTCACAGGTCCCAAGTCTTCCCTCCAGGTGTCTCTAGTTCATGCTGCTGCCAGG 365
QY 92 GlyLeuTrpIleLysLeuProCysThrAspTyrIleGlySerCysThrPheGluHisPhe 111
Db 364 GGGCTTGGGTGCC-----CCATGC----- 344
QY 112 CysAspValLeuAspMet-LeuIleProThrGlyGluPro-----CysPro--GluPro 128
Db 343 -----CTGCCCTCTCCATCTGGGCACCCAGGATCTGCCCTCACACCT 299
QY 129 LeuArgThrTyr--GlyLeuProCysHisCysProPheLysGluGlyThrTyr----- 145
Db 298 CTCCTGGCTGGACAGGTTCACAGGATTCAGGATTCAGGACCTCTGGACCTCTGG 119
QY 146 -----SerLeuProLysSerGluPheAlaValProAspLeuGlu 158
Db 238 CAACCGCTGGGCTCAGGGATTCACAAACAGCTGGTCTCAGCTCCCTCCCTGCT 179
QY 159 LeuProSerTrpLeu-ThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlu 178
Db 178 CTCCTGGTGGCTGCTGCTCAGGTTCACAGGATTCAGGATTCAGGACCTCTGGACCTCTGG 119
QY 178 Y 178
Db 118 C 118

RESULT 9

US-09-201-746-5/c
; Sequence 5, Application US/09201746

[illegible]

```

FEATURE:
NAME/KEY: 3'UTR
LOCATION: 1416..1633
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-097-231-5

Alignment Scores:
Pred. No.: 0.752 Length: 1633
Score: 89.50 Matches: 49
Percent Similarity: 38.25% Conservative: 21
Best Local Similarity: 26.78% Mismatches: 62
Query Match: 8.79% Indels: 51
DB: 3 Gaps: 8

US-10-030-937-9 (1-193) x US-09-097-231-5 (1-1633)

QY 26 LysLysProSerGlnLeu-----SerSer 33
DB 595 AAGAGCCCGTCAGAGATGGACACCTCCAGGACCGGGCTCCTGCTGTGGTGGCAGCCAGC 536
QY 34 PheSerTrpAspAsnCys-----PheGluGlyLysAspProAlaValIleArgSerLeu 51
DB 535 CCCAGCTGGGGGATGGCTGTGGGGGTGGAGTTGAGGGAGCCC-----AGAAAGTCTT 485
QY 52 ThrLeuGluProAspProIleValValProGlyAsnValThrLeuSerValValGlySer 71
DB 484 CTCCTGGGATCCCTGCACAGCCATAGTCTCTCCAGGAAGCAGGAGGAGTCTGTGGAGGC 425
QY 72 ThrSerValProLeuSerSerProLeuLysValAspLeuValLeuGluLysGluValAla 91
DB 424 CTCAGAGTCCCAAGTCTTCCCTCCAGGTGCTCTGCTTGTAGTTTCATGGTCTGCCAGG 365
QY 92 GlyLeuTrpIleLysIleProCysThrAspTyrIleGlySerCysThrPheGluHisPhe 111
DB 364 GGGCTTGGGTGTC---CCATGC-----344
QY 112 CysAspValLeuAspMet-LeuIleProThrGlyGluPro-----CysPro-GluPro 128
DB 343 -----CTGCTCTTCCATCTGGGACCCCGAGATCTGCCCTTCACACCT 299
QY 129 LeuArgThrTyr---GlyLeuProCysHisCysProPheLysGluGlyThrTyr----- 145
DB 298 CTCCTGCTGGCTGGACAGGTGAGCCAGGCGATGCCCGCCAGCCAGTCCAGACACCTCTGGCAT 239
QY 146 -----SerLeuProLysSerGluPheAlaValProAspLeuGlu 158
DB 238 CAACCGCTGGGCTCAGGATCTTCAACAACCAAGCTGTCTCTCAGCTCCCTCCCTGCT 179
QY 159 LeuProSerTrpLeu-ThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlu 178
DB 178 CTTCCTGGTGGCTGCTGCTCAGGTCCAGGACATTCGACACCTCTCGGACCGTCTCTGG 119
QY 178 Y 178
DB 118 C 118

RESULT 11
US-09-353-099-5/c
Sequence 5, Application US/09353099
Patent No. 6476187
GENERAL INFORMATION:
APPLICANT: Cone, Roger D
Fan, Wei
Boston, Bruce A
Kesterton, Robert A
Lu, Dongsi
Chen, Wenbiao
TITLE OF INVENTION: Methods and Reagents for Discovering and
Using Mammalian Melanocortin Receptor Agonists and Antagonists
To Modulate Feeding Behavior in Animals
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

```

```

STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/353,099
FILING DATE: 14-Sep-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/706,281
FILING DATE: 04-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: No. 6476187nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 96,886
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1633 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..461
FEATURE:
NAME/KEY: CDS
LOCATION: 462..1415
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 1416..1633
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-353-099-5

Alignment Scores:
Pred. No.: 0.752 Length: 1633
Score: 89.50 Matches: 49
Percent Similarity: 38.25% Conservative: 21
Best Local Similarity: 26.78% Mismatches: 62
Query Match: 8.79% Indels: 51
DB: 3 Gaps: 8

US-10-030-937-9 (1-193) x US-09-353-099-5 (1-1633)

QY 26 LysLysProSerGlnLeu-----SerSer 33
DB 595 AAGAGCCCGTCAGAGATGGACACCTCCAGGACCGGGCTCCTGCTGTGGTGGCAGCCAGC 536
QY 34 PheSerTrpAspAsnCys-----PheGluGlyLysAspProAlaValIleArgSerLeu 51
DB 535 CCCAGCTGGGGGATGGCTGTGGGGGTGGAGTTGAGGGAGCCC-----AGAAAGTCTT 485
QY 52 ThrLeuGluProAspProIleValValProGlyAsnValThrLeuSerValValGlySer 71
DB 484 CTCCTGGGATCCCTGCACAGCCATAGTCTCTCCAGGAAGCAGGAGGAGTCTGTGGAGGC 425
QY 72 ThrSerValProLeuSerSerProLeuLysValAspLeuValLeuGluLysGluValAla 91
DB 424 CTCAGAGTCCCAAGTCTTCCCTCCAGGTGCTCTGCTTGTAGTTTCATGGTCTGCCAGG 365
QY 92 GlyLeuTrpIleLysIleProCysThrAspTyrIleGlySerCysThrPheGluHisPhe 111
DB 364 GGGCTTGGGTGTC---CCATGC-----344

```



```
Alignment Scores:
Pred. No.: 0.764 Length: 1388
Score: 88.50 Matches: 55
Percent Similarity: 33.63% Conservative: 20
Best Local Similarity: 24.66% Mismatches: 67
Query Match: 8.69% Indels: 81
DB: 4 Gaps: 10

US-10-030-937-9 (1-193) x US-09-976-594-806 (1-1388)

QY 9 LeuLeuLeuAlaLeuLeuLeuLeuAlaThrProAlaGlnAlaHis-----Leu 25
DB 384 GTCCTGATCTCTCGGCTGAGTCTGGCCCGCAGGACCCAGCTGCAGACAGGGACCATC 443
QY 26 LysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPheGluGlyLysAspPro 45
DB 444 CCCAAGCCC-----452
QY 46 AlaValIleArgSerLeuThrLeuGluProAspProIleValValProGlyAsn---Val 64
DB 453 -----ACCTGTGGCTGAGCCAGACTCTGTGATCACCCAGGGGAGTCCCGTC 500
QY 65 ThrLeuSerValValGlySerThrSerValProLeuSerSerProLeuLysValAspLeu 84
DB 501 ACCCTCAGTGTGAGGGGAGCCTTGAAGCCAG-----GAGTACCGTCTA 545
QY 85 ValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThrAsp-----101
DB 546 TATAGGGAGAAATAATCAGCATCT---TGGATTACACGGATACACAGGCGATATGGTGTGAG 602
QY 102 -----TyrIleGlySerCysThrPheGluHis-----110
DB 603 AACGGCCAGTTCACATCCCATCCATCCCTGGGAACACACAGGGCGATATGGTGTGTCAG 662
QY 111 -----PheCysAspValLeuAspMetLeuIle-----119
DB 663 TATTACAGCCGCTGGGTGTCTGAGCTCAGTACGCCCTGGTGTGATGACAGG 722
QY 120 -----ProThrGlyGluProCysProGluProLeuArgThrTyrGly-----133
DB 723 AGTTACCAAAACCCCACTCTCAGCCAGCCCGCTGTGTGTGACCCCAAGAAAGAAC 782
QY 134 -----LeuProCysHisCysProPheLysGluGlyThrTyrSerLeuProLysSerGlu 151
DB 783 GTGACCTGTGTGTGAGTACAGGGGCGAGTTCACACTTCTTCTGACCAAGAGGGG 842
QY 152 PheAlaValProAspLeuGluLeuProSer-----161
DB 843 GCAGGCCATCCCCCACTGCATCTGAGATCAGACGACCAAGCTCAGCAGACCCAGGCTCAA 902
QY 162 -----TrpLeuThrGlyAsnTyrArgIleGluSerValLeu 174
DB 903 TTCGCAATGGTCTGTGACCTCAGCCCGCTGAGGAGCTACAGATGTACAGCTCACTC 962
QY 175 SerSerSer 177
DB 963 AGTCCCAAC 971

RESULT 14
US-09-785-381-6/c
; Sequence 6, Application US/09785381
; Patent No. 6602992
; GENERAL INFORMATION:
; APPLICANT: DALLOS, Peter
; APPLICANT: ZHENG, Jing
; APPLICANT: MADISON, Laird
; TITLE OF INVENTION: A MAMMALIAN PRESTIN
; FILE REFERENCE: 0290-37UJ
; CURRENT APPLICATION NUMBER: US/09/785,381
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,461
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 16

US-09-785-381-6 (1-29485)
; SEQ ID NO 6
; LENGTH: 29485
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-785-381-6
Alignment Scores:
Pred. No.: 141 Length: 29485
Score: 88.00 Matches: 47
Percent Similarity: 38.01% Conservative: 18
Best Local Similarity: 27.49% Mismatches: 63
Query Match: 8.64% Indels: 43
DB: 4 Gaps: 9

US-10-030-937-9 (1-193) x US-09-785-381-6 (1-29485)

QY 20 ProAlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCys 39
DB 12973 CCTCCCAAAACCAAGCTCGAGCATCCAGTTCGACCTCA-----GACTGTCTGT 12926
QY 40 PheGluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleVal 59
DB 12925 -----GTTGGCAGCGAAGATTTCAAGCCA-----12902
QY 60 ValProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerPro 79
DB 12901 -----GTGAATCTTACCTTGTCTGGCTCCGTTGGGGTGGGACTCACTGAGCCA 12854
QY 80 LeuLysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCys 99
DB 12853 GGCACCGAAGGAATGCTGCTGCTGCCAGTTGCAAAAGACTGTGGGAAAGCGCAGTATC 12794
QY 100 ThrAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMet-Leu-- 118
DB 12793 TGGGCCAGATTG---CACTGTCTCCAGCAGCAGATTGCTCAGCGCTCCCTTGGCTAGG 12737
QY 119 -----IleProThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuPr 135
DB 12736 AAGGGAAATCCCGCAGC-----CCTTGTGGTCTGCCACACTCTGCTTCACTCACC 12683
QY 135 0-----CysHisCysProPheLysGluGlyThrTyrSerLeuProLysSerGlu 151
DB 12682 CTCATGCGCTGCACCCCACTGTCCA-----ACCAGTCCCAATGAGAT 12641
QY 151 uPheAlaValProAspLeuGluLeuProSerTrpLeuThr-----ThrGlu 166
DB 12640 GAACCGGTTCTCCTCAGTTGGAAATGCCGAAATCACCACCTTCTGCATGTGCTGTGG 12581
QY 166 YAsnTyrArgIleGluSerValLeuSerSer 176
DB 12580 GAGCTGCAGACTGGAGCTGTCTCTATTACAGC 12550

RESULT 15
US-09-735-934A-3
; Sequence 3, Application US/09735934A
; Patent No. 6372468
; GENERAL INFORMATION:
; APPLICANT: LI, Jiayin et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000851
; CURRENT APPLICATION NUMBER: US/09/735,934A
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-934A-3
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 22:22:21 ; Search time 887.8 Seconds
(without alignments)
1174.701 Million cell updates/sec

Title: US-10-030-937-9
Perfect score: 1018
Sequence: 1 MQSLMQAPLLIALGLLALP.....LSSGKRLGCIKIAASLKI 193

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3627988 seqs, 2701811610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlip
-Q=/cgn2_1/USPTO_spool_p/US10030937/runat_16112004_153015_3035/app_query.fasta_1.789
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct-THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10030937@cgn_1_1223_brunat_16112004_153015_3035
-NCFU=6 -ICPU=3 -NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1005	98.7	2436	9	US-09-954-531-380	Sequence 380, App
2	1005	98.7	2436	10	US-09-525-978B-81	Sequence 81, Appl
3	1005	98.7	2478	15	US-10-170-385-390	Sequence 390, App
4	1000	98.2	1935	10	US-09-971-392-102	Sequence 102, App
5	1000	98.2	2384	9	US-09-822-849A-53	Sequence 53, Appl
6	740.5	72.7	1983	16	US-10-388-934-167	Sequence 167, App
7	424	41.8	577	16	US-10-264-049-436	Sequence 436, App
c	354	34.7	475	9	US-09-864-761-1518	Sequence 1518, App
9	333	32.7	448	11	US-09-969-034-4215	Sequence 4215, App
10	272	26.7	546	13	US-10-027-632-207798	Sequence 207798, App
11	272	26.7	546	13	US-10-027-632-207799	Sequence 207799, App
12	272	26.7	546	13	US-10-027-632-207800	Sequence 207800, App
13	272	26.7	546	13	US-10-027-632-207801	Sequence 207801, App
14	272	26.7	546	15	US-10-027-632-207798	Sequence 207798, App
15	272	26.7	546	15	US-10-027-632-207799	Sequence 207799, App
16	272	26.7	546	15	US-10-027-632-207800	Sequence 207800, App
17	272	26.7	546	15	US-10-027-632-207801	Sequence 207801, App
c	270	26.5	145	9	US-09-864-761-18277	Sequence 18277, A
18	262.5	25.8	250000	15	US-10-225-810-26	Sequence 26, Appl
19	254	25.0	857	13	US-10-027-632-164063	Sequence 164063, App
20	254	25.0	857	13	US-10-027-632-164064	Sequence 164064, App
21	254	25.0	857	13	US-10-027-632-164065	Sequence 164065, App
22	254	25.0	857	13	US-10-027-632-164066	Sequence 164066, App
23	254	25.0	857	15	US-10-027-632-164063	Sequence 164063, App
24	254	25.0	857	15	US-10-027-632-164064	Sequence 164064, App
25	254	25.0	857	15	US-10-027-632-164065	Sequence 164065, App
26	221	21.7	380	10	US-09-764-891-2290	Sequence 2290, App
27	145	14.2	593	13	US-10-027-632-277778	Sequence 277778, App
28	145	14.2	593	13	US-10-027-632-277779	Sequence 277779, App
c	103	10.1	4821	18	US-10-425-115-178335	Sequence 178335, App
29	98.5	9.7	819	18	US-10-425-115-112475	Sequence 112475, App
30	97.5	9.6	20966	9	US-09-776-976-7	Sequence 7, Appl
31	97.5	9.6	20966	9	US-09-758-055-7	Sequence 7, Appl
32	97.5	9.6	20966	9	US-09-909-547-7	Sequence 7, Appl
33	97.5	9.6	20966	15	US-10-231-814-7	Sequence 1, Appl
34	97.5	9.6	20966	15	US-10-376-460-1	Sequence 1, Appl
35	97.5	9.6	20966	15	US-10-285-833-7	Sequence 7, Appl
36	97.5	9.6	20966	16	US-10-027-632-173933	Sequence 173933, App
c	93	9.1	727	13	US-10-027-632-173934	Sequence 173934, App
37	93	9.1	727	13	US-10-027-632-173935	Sequence 173935, App
38	93	9.1	727	13	US-10-027-632-173936	Sequence 173936, App
39	93	9.1	727	13	US-10-027-632-173937	Sequence 173937, App
40	93	9.1	727	13	US-10-027-632-173938	Sequence 173938, App
41	93	9.1	727	15	US-10-027-632-173939	Sequence 173939, App
42	93	9.1	727	15	US-10-027-632-173940	Sequence 173940, App
43	93	9.1	727	15	US-10-027-632-173941	Sequence 173941, App
c	93	9.1	727	15	US-10-027-632-173942	Sequence 173942, App
44	93	9.1	727	15	US-10-027-632-173943	Sequence 173943, App
45	93	9.1	2143	16	US-10-108-260A-141	Sequence 141, App

ALIGNMENTS

RESULT 1
US-09-954-531-380
; Sequence 380, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cancer
; TITLE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22

Fri Nov 19 14:12:52 2004

us-10-030-937-9.p2n.rnpb

```

; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 380
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-531-380

Alignment Scores:
Pred. No.: 2,33e-117 Length: 2436
Score: 1005.00 Matches: 191
Percent Similarity: 98.96% Conservative: 0
Best Local Similarity: 98.96% Mismatches: 2
Query Match: 98.72% Indels: 2
DB: 9 Gaps: 0

US-10-030-937-9 (1-193) x US-09-531-380 (1-2436)
QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db 59 ATGCAGTCCCTGATGAGGCTCCCTCTGATCGCCCTGGGCTTGTCTTCGCGACCCCT 118
QY 21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
Db 119 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTCTCTGGGATACTGTGAT 178
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 179 GAAGGAGAGACCTCGGCTGATCAGAGCTGACTCTGGAGCTGACCCCATCGTCTG 238
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerProLeu 80
Db 239 CCTGGAATGTGACCTCAGTGTGGGAGGAGTGGCTCTGGATCAAGATCCCATGCACA 298
QY 81 LysValAspLeuValLeuGlyLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
Db 299 AAGTGGATTAGTTTGGAGAGGAGTGGCTCTGGATCAAGATCCCATGCACA 358
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 359 GACTACATTGGCAGCTGACTTTGAACACTTCTGTGATGTCTTGACATGTTAATTCCT 418
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 419 ACTGGGAGCCCTGCCAGAGCCCTCGTACCTATGGGCTTCTTGCCACTGTCCCTTC 478
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluPro 160
Db 479 AAAGAAGAACCTACTCACTGCCCAAGAGCGAATTCGTTGTGCTGACCTGGAGTGCCT 538
QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
Db 539 AGTTGGCTACCACCGGAACTACCGCATAGAGAGCGTCTCTGAGCAGCAGTGGGAAGCGT 598
QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 599 CTGGGCTGCATCAAGATCGCTGCTCTCTAAAGGGCATA 637

RESULT 2
US-09-525-978B-81
; Sequence 81, Application US/09525978B
; Publication No. US20030049722A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Caras, Ingrid W.
; APPLICANT: Hevezi, Peter
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING MACROPHAGE DEVELOPMENT
; TITLE OF INVENTION: RELATED DISORDERS, COMPOSITIONS, AND METHODS OF
; TITLE OF INVENTION: SCREENING FOR MACROPHAGE DEVELOPMENT MODULATORS
; FILE REFERENCE: A-67413-1/DJB/JJD
; CURRENT APPLICATION NUMBER: US/09/525,978B
; CURRENT FILING DATE: 2000-03-15

; PRIOR APPLICATION NUMBER: USSN 60/124,530
; PRIOR FILING DATE: 1999-03-15
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-525-978B-81

Alignment Scores:
Pred. No.: 2,33e-117 Length: 2436
Score: 1005.00 Matches: 191
Percent Similarity: 98.96% Conservative: 0
Best Local Similarity: 98.96% Mismatches: 2
Query Match: 98.72% Indels: 2
DB: 10 Gaps: 0

US-10-030-937-9 (1-193) x US-09-525-978B-81 (1-2436)
QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db 59 ATGCAGTCCCTGATGAGGCTCCCTCTGATCGCCCTGGGCTTGTCTTCGCGACCCCT 118
QY 21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
Db 119 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTCTCTGGGATACTGTGAT 178
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 179 GAAGGAGAGACCTCGGCTGATCAGAGCTGACTCTGGAGCTGACCCCATCGTCTG 238
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerProLeu 80
Db 239 CCTGGAATGTGACCTCAGTGTGGGAGGAGTGGCTCTGGATCAAGATCCCATGCACA 298
QY 81 LysValAspLeuValLeuGlyLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
Db 299 AAGTGGATTAGTTTGGAGAGGAGTGGCTCTGGATCAAGATCCCATGCACA 358
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 359 GACTACATTGGCAGCTGACTTTGAACACTTCTGTGATGTCTTGACATGTTAATTCCT 418
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 419 ACTGGGAGCCCTGCCAGAGCCCTCGTACCTATGGGCTTCTTGCCACTGTCCCTTC 478
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluPro 160
Db 479 AAAGAAGAACCTACTCACTGCCCAAGAGCGAATTCGTTGTGCTGACCTGGAGTGCCT 538
QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
Db 539 AGTTGGCTACCACCGGAACTACCGCATAGAGAGCGTCTCTGAGCAGCAGTGGGAAGCGT 598
QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 599 CTGGGCTGCATCAAGATCGCTGCTCTCTAAAGGGCATA 637

RESULT 3
US-10-170-385-390
; Sequence 390, Application US/10170385
; Publication No. US20030203372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart

```

APPLICANT: Kingsman, Susan Mary
APPLICANT: Krige, David
TITLE OF INVENTION: ANALYSIS METHOD
FILE REFERENCE: 53268200100
CURRENT APPLICATION NUMBER: US/10/170,385
CURRENT FILING DATE: 2002-06-12
PRIOR APPLICATION NUMBER: PCT/GB02/01662
PRIOR FILING DATE: 2002-04-08
PRIOR APPLICATION NUMBER: PCT/GB01/05458
PRIOR FILING DATE: 2001-12-10
NUMBER OF SEQ ID NOS: 549
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 390
LENGTH: 2478
TYPE: DNA
ORGANISM: Homo Sapiens
US-10-170-385-390

Alignment Scores:

Pred. No.: 2,39e-117 Length: 2478
Score: 1005.00 Matches: 191
Percent Similarity: 98.96% Conservative: 0
Best Local Similarity: 98.96% Mismatches: 2
Query Match: 98.72% Indels: 0
DB: 15 Gaps: 0

US-10-030-937-9 (1-193) x US-10-170-385-390 (1-2478)

Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db 96 ATGCAGTCCCTGATGCAGGCTCCCTCTGATCGCCCTGGGCTGTTCTCGCGACCCCT 155
Qy 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerThrAspAsnCysPhe 40
Db 156 GCGCAAGCCACCTGAAAGGCCATCCAGCTCAGTAGCTTTCTCGGATAAATGTGAT 215
Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 216 GAAGGAGAGGACCTCGCGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTCGT 275
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 276 CCTGGAATGTGACCTCAGTGTCTGGGAGCAGCAGGCTGTCCTCTGAGTTCTCTCTG 335
Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrIleLysIleProCysThr 100
Db 336 AAGGTGGATTTAGTTTGGAGAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGCACA 395
Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 396 GACTACATTTGCAGCTGACCTTTTGAACACTTCTGTGATGTGTGACATGTTAATTCCT 455
Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 456 ACTGGGAGCCCTGCCAGAGCCCTCGGTACCTATGGGCTTCTTGCCACTGTCCTTC 515
Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
Db 516 AAAGAAGGAACCTACTCACTGCCCAAGAGCGAATTCGTTGTGCTGACCTGGAGTGGCC 575
Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
Db 576 AGTTGGCTCACCCCGGAACCTACCGCATAGAGAGCGTCTCGAGCAGCGTGGGAAGCGT 635
Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 636 CTGGGCTGCATCAGATCGCTGCTCTCTAAAGGGGCATA 674

RESULT 4

US-09-971-392-102
; Sequence 102, Application US/09971392
; Publication No. US20030134283A1
; GENERAL INFORMATION:

APPLICANT: Peterson, David P.
APPLICANT: Pearson, Cecelia I.
APPLICANT: Cocks, Benjamin G.
TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
FILE REFERENCE: PA-0029 US
CURRENT APPLICATION NUMBER: US/09/971,392
CURRENT FILING DATE: 2001-10-03
PRIOR APPLICATION NUMBER: 60/237,652
PRIOR FILING DATE: 2000-10-03
NUMBER OF SEQ ID NOS: 260
SOFTWARE: PERL Program
SEQ ID NO 102
LENGTH: 1935
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Template ID: 977615.8
US-09-971-392-102

Alignment Scores:

Pred. No.: 7,28e-117 Length: 1935
Score: 1000.00 Matches: 190
Percent Similarity: 98.45% Conservative: 0
Best Local Similarity: 98.45% Mismatches: 3
Query Match: 98.23% Indels: 0
DB: 10 Gaps: 0

US-10-030-937-9 (1-193) x US-09-971-392-102 (1-1935)

Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db 102 ATGCAGTCCCTGATGCAGGCTCCCTCTGATCGCCCTGGGCTGTTCTCGCGCCCT 161
Qy 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerThrAspAsnCysPhe 40
Db 162 GCGCAAGCCACCTGAAAGGCCATCCAGCTCAGTAGCTTTCTGGGATAAATGTGAT 221
Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 222 GAAGGAGAGGACCTCGCGTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCGTCGT 281
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 282 CCTGGAATGTGACCTCAGTGTCTGGGAGCAGCAGGCTGTCCTCTGAGTTCTCTCTG 341
Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrIleLysIleProCysThr 100
Db 342 AAGGTGGATTTAGTTTGGAGAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGCACA 401
Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 402 GACTACATTTGCAGCTGACCTTTTGAACACTTCTGTGATGTGTGACATGTTAATTCCT 461
Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 462 ACTGGGAGCCCTGCCAGAGCCCTCGGTACCTATGGGCTTCTTGCCACTGTCCTTC 521
Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
Db 522 AAAGAAGGAACCTACTCACTGCCCAAGAGCGAATTCGTTGTGCTGACCTGGAGTGGCC 581
Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
Db 582 AGTTGGCTCACCCCGGAACCTACCGCATAGAGAGCGTCTCGAGCAGCGTGGGAAGCGT 641
Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 642 CTGGGCTGCATCAGATCGCTGCTCTCTAAAGGGGCATA 680

RESULT 5

US-09-971-392-102
; Sequence 102, Application US/09971392
; Publication No. US20030134283A1
; GENERAL INFORMATION:

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99
 US-09-864-761-1518

Alignment Scores:
 Pred. No.: 2,4e-35 Length: 475
 Score: 354.00 Matches: 61
 Percent Similarity: 100.00% Conservatives: 1
 Best Local Similarity: 98.39% Mismatches: 0
 Query Match: 34.77% Indels: 0
 DB: 9 Gaps: 0

US-10-030-937-9 (1-193) x US-09-864-761-1518 (1-475)

QY 81 LysValAspLeuValLeuGluValAlaGlyLeuTrpLeuLysValLeuProCysThr 100
 Db 468 CAGGTGGATTAGTTTGGAGAGAGAGGTGGCTCTGGATCAGATCCCATGCACA 409
 QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 408 GACTACATTGGCAGCTGACCTTTGAACACTTCTGTGATGTGCTTGACATGTTAATTCCT 349
 QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 348 ACTGGGAGCCCTGCCAGAGCCCTGCGTACCTATGGGCTTCTTGCCACTGTCCTTC 289
 QY 141 LysGlu 142
 Db 288 AAAGAA 283

RESULT 9

US-09-969-034-4215
 ; Sequence 4215, Application US/09969034
 ; Publication No. US20040110668A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Burgess, Christopher C.
 ; APPLICANT: Astle, Jon H.
 ; APPLICANT: Carroll, Eddie III
 ; APPLICANT: Catino, Theodore J.
 ; APPLICANT: Dwivedi, Poornima
 ; APPLICANT: Molino, Gary A.
 ; APPLICANT: Thiagalingam, Arunthathi
 ; APPLICANT: Lewis, Marcia E.
 ; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
 ; TITLE OF INVENTION: Expressed in Cancer Tissue
 ; FILE REFERENCE: 1657/1032
 ; CURRENT APPLICATION NUMBER: US/09/969,034
 ; CURRENT FILING DATE: 2001-10-02
 ; PRIOR APPLICATION NUMBER: 60/237,271
 ; PRIOR FILING DATE: 2000-02-10
 ; NUMBER OF SEQ ID NOS: 4494
 ; SOFTWARE: Fast-Seq for Windows Version 4.0
 ; SEQ ID NO 4215
 ; LENGTH: 448
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: 241, 277, 288, 295, 299, 300, 304, 310, 316, 343, 346, 356,
 ; LOCATION: 364, 370, 396, 397, 406, 410, 415, 424, 437
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-969-034-4215

Alignment Scores:
 Pred. No.: 1.05e-32 Length: 448
 Score: 333.00 Matches: 62
 Percent Similarity: 98.41% Conservatives: 0
 Best Local Similarity: 98.41% Mismatches: 1
 Query Match: 32.71% Indels: 0
 DB: 11 Gaps: 0

US-10-030-937-9 (1-193) x US-09-969-034-4215 (1-448)

QY 131 ThrTyrGlyLeuProCysHisCysProPheLysGluGlyThrTyrSerLeuProLysSer 150
 Db 1 ACCTATGGGCTTCCTTGGCCACTGTCTCTTCAAGAGGAACCTTACTTACCTGCCAAGAGC 60
 QY 151 GluPheAlaValProAspLeuGluLeuProSerTrpLeuThrThrGlyAsnTyrArgIle 170
 Db 61 GAATTCGTTGTGCTGACCTGGAGCTGCCAGTTGGTTCACCCGGAATGTACCGCATA 120
 QY 171 GluSerValLeuSerSerGlyLysArgLeuGlyCysIleLysIleAlaAlaSerLeu 190
 Db 121 GAGAGCGTCTGTGAGCAGCAGTGGGAAGCGTCTGGGCTGCATCAAGATCGCTCTCTTA 180
 QY 191 LysGlyIle 193
 Db 181 AAGGGCATA 189

RESULT 10

US-10-027-632-207798
 ; Sequence 207798, Application US/10027632
 ; Publication No. US20020198371A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; TITLE OF INVENTION: Polymorphisms in the Human Genome
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; CURRENT FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 207798
 ; LENGTH: 546
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-027-632-207798

Alignment Scores:
 Pred. No.: 8.42e-25 Length: 546
 Score: 272.00 Matches: 57
 Percent Similarity: 86.76% Conservatives: 2
 Best Local Similarity: 83.82% Mismatches: 7
 Query Match: 26.72% Indels: 2
 DB: 13 Gaps: 1

US-10-030-937-9 (1-193) x US-10-027-632-207798 (1-546)

QY 27 LysProSerGlnLeuSerSerPheSerTrpAspAsnCysPheGluGlyLysAspProAla 46
 Db 255 CAGCCATCCAGCTCAGTAGCTTTCTTGGGTAACGTGATGAAGGAAGGACCCCTGCG 314
 QY 47 ValIleArgSerLeuThrLeuGluProAspProIleValValProGlyAsnValThrLeu 66
 Db 315 GTGATCAGAGCTGACTCTGGAGCCTGACCCCATCTCTCTTCTGGAAATGTGACCTC 374
 QY 67 SerValValGlySerThrSerValProLeuSerSerProLeuLysValAspLeuValLeu 86
 Db 375 AGTGTCRTGGGAGCAGCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 434

```
QY      87  -----GluLysGluValAlaGly 92
Db      435  GGTGAGAGGAGGAGGTGCGAGGG 458

RESULT 11
US-10-027-632-207799
; Sequence 207799, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207799
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-207799

Alignment Scores:
Pred. No.:      8,42e-25      Length:      546
Score:          272.00      Matches:      57
Percent Similarity: 86.76%      Conservative: 2
Best Local Similarity: 83.82%      Mismatches: 7
Query Match:    26.72%      Indels:      2
DB:             13          Gaps:        1

US-10-030-937-9 (1-193) x US-10-027-632-207799 (1-546)
QY      27  LysProSerGlnLeuSerSerPheSerTrpAspAsnCysPheGluGlyAspProAla 46
Db      255  CAGCCATCCAGCTCAGTAGCTTTCTCTGGGATACTGTGATGAAGGAGGAGCCCTGCG 314

QY      47  ValIleArgSerLeuThrLeuGluProAspProIleValValProGlyAsnValThrLeu 66
Db      315  GTGATCAGAGCGCTGACTCTGGAGCGCTGACCCCATCTCTCTGGAAATGTGACCCCTC 374

QY      67  SerValValGlySerThrSerValProLeuSerSerProLeuLysValAspLeuValLeu 86
Db      375  AGTGTCTGGGCGAGCACCAGTGTCCCTCTGAGGTCTCTCTGAGGTGAGCCTGGGGGTG 434

QY      87  -----GluLysGluValAlaGly 92
Db      435  GGTGAGAGGAGGAGGTGCGAGGG 458

RESULT 12
US-10-027-632-207800
; Sequence 207800, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207799
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-207799

Alignment Scores:
Pred. No.:      8,42e-25      Length:      546
Score:          272.00      Matches:      57
Percent Similarity: 86.76%      Conservative: 2
Best Local Similarity: 83.82%      Mismatches: 7
Query Match:    26.72%      Indels:      2
DB:             13          Gaps:        1

US-10-030-937-9 (1-193) x US-10-027-632-207799 (1-546)
QY      27  LysProSerGlnLeuSerSerPheSerTrpAspAsnCysPheGluGlyAspProAla 46
Db      255  CAGCCATCCAGCTCAGTAGCTTTCTCTGGGATACTGTGATGAAGGAGGAGCCCTGCG 314

QY      47  ValIleArgSerLeuThrLeuGluProAspProIleValValProGlyAsnValThrLeu 66
Db      315  GTGATCAGAGCGCTGACTCTGGAGCGCTGACCCCATCTCTCTGGAAATGTGACCCCTC 374

QY      67  SerValValGlySerThrSerValProLeuSerSerProLeuLysValAspLeuValLeu 86
Db      375  AGTGTCTGGGCGAGCACCAGTGTCCCTCTGAGGTCTCTCTGAGGTGAGCCTGGGGGTG 434

QY      87  -----GluLysGluValAlaGly 92
Db      435  GGTGAGAGGAGGAGGTGCGAGGG 458

RESULT 13
US-10-027-632-207801
; Sequence 207801, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
```

; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207801
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-207801

Alignment Scores:
Pred. No.: 8.42e-25 Length: 546
Score: 272.00 Matches: 57
Percent Similarity: 86.76% Conservative: 2
Best Local Similarity: 83.82% Mismatches: 7
Query Match: 26.72% Indels: 2
DB: 13 Gaps: 1

US-10-030-937-9 (1-193) x US-10-027-632-207801 (1-546)

Qy 27 LysProSerGlnLeuSerSerPheSerTrpAspAsnCysPheGluGlyLysAspProAla 46
Db 255 CAGCCATCCAGCTCAGTAGCTTTCTGGGATAACTGTGATGAAGGAAGGACCCCTGCG 314
Qy 47 ValIleArgSerLeuThrLeuGluProAspProIleValValProGlyAsnValThrLeu 66
Db 315 GTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCTCGTTCTCTGGAATGTGACCCCTC 374
Qy 67 SerValValGlySerThrSerValProLeuSerSerProLeuLysValAspLeuValLeu 86
Db 375 AGTGTCRTGGCGACGACGAGTGTCCCTGAGTTCTCTCTGAAGGTGAGCCTGGGGGTG 434
Qy 87 -----GluLysGluValAlaGly 92
Db 435 GGTGGAGAAGGGAGGTGCGAGGG 458

RESULT 14

US-10-027-632-207798
; Sequence 207798, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207798
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human

US-10-027-632-207798

Alignment Scores:
Pred. No.: 8.42e-25 Length: 546
Score: 272.00 Matches: 57
Percent Similarity: 86.76% Conservative: 2

Best Local Similarity: 83.82% Mismatches: 7
Query Match: 26.72% Indels: 2
DB: 15 Gaps: 1

US-10-030-937-9 (1-193) x US-10-027-632-207798 (1-546)

Qy 27 LysProSerGlnLeuSerSerPheSerTrpAspAsnCysPheGluGlyLysAspProAla 46
Db 255 CAGCCATCCAGCTCAGTAGCTTTCTGGGATAACTGTGATGAAGGAAGGACCCCTGCG 314
Qy 47 ValIleArgSerLeuThrLeuGluProAspProIleValValProGlyAsnValThrLeu 66
Db 315 GTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCTCGTTCTCTGGAATGTGACCCCTC 374
Qy 67 SerValValGlySerThrSerValProLeuSerSerProLeuLysValAspLeuValLeu 86
Db 375 AGTGTCRTGGCGACGACGAGTGTCCCTGAGTTCTCTCTGAAGGTGAGCCTGGGGGTG 434
Qy 87 -----GluLysGluValAlaGly 92
Db 435 GGTGGAGAAGGGAGGTGCGAGGG 458

RESULT 15

US-10-027-632-207799
; Sequence 207799, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207799
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human

Alignment Scores:
Pred. No.: 8.42e-25 Length: 546
Score: 272.00 Matches: 57
Percent Similarity: 86.76% Conservative: 2
Best Local Similarity: 83.82% Mismatches: 7
Query Match: 26.72% Indels: 2
DB: 15 Gaps: 1

US-10-030-937-9 (1-193) x US-10-027-632-207799 (1-546)

Qy 27 LysProSerGlnLeuSerSerPheSerTrpAspAsnCysPheGluGlyLysAspProAla 46
Db 255 CAGCCATCCAGCTCAGTAGCTTTCTGGGATAACTGTGATGAAGGAAGGACCCCTGCG 314
Qy 47 ValIleArgSerLeuThrLeuGluProAspProIleValValProGlyAsnValThrLeu 66
Db 315 GTGATCAGAAGCCTGACTCTGGAGCCTGACCCCATCTCGTTCTCTGGAATGTGACCCCTC 374

Qy 67 SerValValGlySerThrSerValProLeuSerSerProLeuLysValAspLeuValLeu 86
Db 375 AGTGTCTGTGGGCAGCACCGAGTGTCCCCCTGAGTTCTCCTCTGAGGTTGAGCCTGGGGGTG 434
Qy 87 -----GluLysGluValAlaGly 92
Db 435 GGTGGAGAGGGGAGGTGCGAGGG 458

Search completed: November 19, 2004, 03:29:39
Job time : 892.8 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 19:16:51 ; Search time 6209.45 Seconds
(without alignments)
1132.607 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 1018

Sequence: 1 MQSLMQAPLLALGILLATP.....LSSGKRLGCIKIASLKI 193

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO_spool_p/US10030937/runat_16112004_153014_2925/app_query.fasta_1.789
-DB=EST -QFMT=FASTAP -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40 cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937@cgn_1_1_9321@runat_16112004_153014_2925 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gsi1:*
9: gb_gsi2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1000	98.2	698	2 BF509172	BF509172 UI-H-B14-
2	1000	98.2	937	5 BUI151364	BUI151364 AGENCOURT
3	999	98.1	858	4 BI820051	BI820051 603037236
4	996	97.8	2338	3 CR626644	CR626644 full-length
5	994	97.6	948	5 BUI556006	BUI556006 AGENCOURT
6	989	97.2	909	5 BQ643369	BQ643369 AGENCOURT
7	989	97.2	973	4 BM474816	BM474816 AGENCOURT
8	986	96.9	651	4 BM723945	BM723945 UI-E-E01-
9	983	96.6	726	4 BG762599	BG762599 602734472

10	982	96.5	784	4 BI838554	BI838554 603086219
11	982	96.5	949	5 BQ600062	BQ600062 AGENCOURT
12	981	96.4	784	4 BG323734	BG323734 602421833
13	977	96.0	733	4 BI856212	BI856212 603382936
14	977	96.0	842	4 BI091220	BI091220 602856051
15	977	96.0	912	1 AL543858	AL543858 602856051
16	977	96.0	994	1 AL548441	AL548441 602856051
17	977	96.0	1060	1 AL550565	AL550565 602856051
18	973	95.6	910	1 AL560604	AL560604 602856051
19	965	94.8	997	4 BG830059	BG830059 602764494
20	962	94.5	863	4 BG478588	BG478588 602524087
21	961	94.4	813	4 BG913328	BG913328 602812047
22	959	94.2	760	4 BG770447	BG770447 602734356
23	954	93.7	1018	2 BE735010	BE735010 601567832
24	946	92.9	1138	1 AL513584	AL513584 601567832
25	941	92.4	818	1 AL552056	AL552056 601567832
26	940	92.3	847	5 BQ220522	BQ220522 AGENCOURT
27	939	92.2	1067	2 BF528447	BF528447 602043611
28	926	91.0	870	5 BX404478	BX404478 602043611
29	919	90.3	711	4 BG479322	BG479322 602526351
30	919	90.3	1031	4 BM476220	BM476220 AGENCOURT
31	916	89.9	643	1 AU135438	AU135438 601567832
32	915	89.9	997	4 BM561693	BM561693 AGENCOURT
33	913	89.7	559	1 AU280628	AU280628 60280628
34	903	88.7	705	4 BG912821	BG912821 602807283
35	894	87.8	885	6 CA453926	CA453926 AGENCOURT
36	893	87.7	601	5 BX506263	BX506263 DKFZP686F
37	887	87.1	767	4 BI161208	BI161208 602865666
38	861	84.6	895	5 BQ676659	BQ676659 AGENCOURT
39	856	84.1	1130	2 BE613752	BE613752 601504554
40	850	83.5	646	6 CD703135	CD703135 EST19726
41	846	83.1	911	2 BF127825	BF127825 601810539
42	831	81.6	554	4 BG830178	BG830178 602764754
43	819	80.5	611	4 BG623044	BG623044 602647926
44	802	78.8	476	4 BM147068	BM147068 TCAAP1Q10
45	796	78.2	603	2 BF9555424	BF9555424 MR4-NN119

ALIGNMENTS

RESULT 1
BF509172
LOCUS UI-H-B14-av-c-10-0-UI-sl NCI CGAP_Sub8 Homo sapiens cdna clone
DEFINITION IMAGE:3086203 3', mRNA sequence.
ACCESSION BF509172.1 GI:11592470
VERSION BF509172.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1. (bases 1 to 698)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Unpublished (1997)
Tumor Gene Index
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
is likely internal to the message. CDNA Library Preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M3 Forward
POLYA-No. Location/Qualifiers
1. .698
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3086203"
/lab_host="DHI08 (Life Technologies)"

/clone_lib="NCI_CGAP_Sub8"
/notes=vector: pT73D-Fac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; NCI CGAP Sub8
is a subtracted library derived from NCI CGAP Subs. The
NCI CGAP Sub8 library had 2.5 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub8 was used
as a tracer in a subtractive hybridization with a driver
comprising: a pool of clones from NCI_CGAP_Sub5 (IMAGE
clone ids 2732833-2737415, 3068040-3069191; 25% of the
driver population), a pool of clones from NCI_CGAP_Sub4
(IMAGE clone ids 2723592-2729326; 25% of the driver
population), NCI_CGAP_Sub6 (pool AIP-AJU, IMAGE ids
2728969-2733190; 25% of the driver population), and
NCI_CGAP_Sub7 (IMAGE ids 3069192-3072238,
3081864-3084550; 25% of the driver population).
Subtraction was performed as previously described
[Bonaldo, Lennon & Soares (1996): Normalization and
Genome Research 6, 791-806.
TAG_SEQ=None found"

ORIGIN

Alignment Scores:
Pred. No.: 2,44e-91 Length: 698
Score: 1000.00 Matches: 190
Percent Similarity: 98.45% Conservative: 0
Best Local Similarity: 98.45% Mismatches: 3
Query Match: 98.23% Indels: 0
DB: 2 Gaps: 0

US-10-030-937-9 (1-193) x BF509172 (1-698)

Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuLeuAlaThrPro 20
Db 50 ATGCAGTCCCTGATGAGGCTCCCTCTGATGCGCCCTGGCTTCTCGCGGCCCT 109
Qy 21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerThrPheAsnCysPhe 40
Db 110 GCGCAAGCCACCTCGAAAGGCCATCCAGCTCAGTAGTCTTCTCGGATGATGAT 169
Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 170 GAAGGAGGACCTCGGGTATCAGAGGCTGACTCTGGAGCTGACCCATCGTCTT 229
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 230 CCTGGAATGTGACCTCAGTGTCTGGGAGCAGCAGTGTCTGCTGCTCTCTCTG 289
Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrIleValIleProCysThr 100
Db 290 AAGGTGATATTAGTTTGGAGAGGAGGTGGCTGGCTCTGATCAAGATCCCATGACA 349
Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 350 GACTACATGGCAGCTGACTTGAACACTCTGTGTGTGTGTGTGTGTGTGTGTGT 409
Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 410 ACTGGGAGCCCTCCAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 469
Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
Db 470 AAAGAGGAACCTACTACTGCCCCAAGAGCGAATTCGTGTGTGTGTGTGTGTGT 529
Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
Db 530 AGTTGGCTCACACCGGAACTACCGCATAGAGCGTCTCTGAGCAGCAGTGGAGCGT 589
Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 590 CTGGGCTGCATCAAGATCGTGCCTCTCTAAAGGGCATA 628

RESULT 2

BU151364
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BU151364
AGENCY: 8119275
Lupski dorsal root ganglion Homo sapiens cdna
clone IMAGE:6179622 5', mRNA sequence.
BU151364
EST.
BU151364.1 GI:22664896
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 937)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@nih.gov
Tissue Procurement: Dr. James R. Lupski
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13561 Row: b Column: 07
High quality sequence stop: 684.

FEATURES

source
1. 937
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6179622"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/clone_lib="Lupski dorsal root ganglion"
/note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGCTCCG-3' and
5'-GACTAGTCTAGTCGCGAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."

ORIGIN

Alignment Scores:
Pred. No.: 3,78e-91 Length: 937
Score: 1000.00 Matches: 190
Percent Similarity: 98.45% Conservative: 0
Best Local Similarity: 98.45% Mismatches: 3
Query Match: 98.23% Indels: 0
DB: 5 Gaps: 0

US-10-030-937-9 (1-193) x BU151364 (1-937)

Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db 74 ATGCAGTCCCTGATGAGGCTCCCTCTGATGCGCCCTGGCTTCTCGCGGCCCT 133
Qy 21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerThrPheAsnCysPhe 40
Db 134 GCGCAAGCCACCTCGAAAGGCCATCCAGCTCAGTAGTCTTCTCGGATGATGAT 193
Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 194 GAAGGAGGAGGACCTCGGGTATCAGAGCCTGACTCTGGAGCCTGACCCATCGTCT 253
Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80

Db 254 CCTGGAATGTGACCTCAGTGTGCTGGCAGCACCAGTGTCCCCCTGAGTTCTCCTCTG 313
 Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPheLysIleProCysThr 100
 Db 314 AAGGTGGAATTTAGTTTGGAGAAGAGGTGGCTGCTGGATCAAGATCCCATGCACA 373
 Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 374 GACTACATGGCAGCTGTACCTTTGAACACTCTCTGTGATGCTGTGACATGTTAATTCCT 433
 Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 434 ACTGGGAGCCCTGCCAGAGCCCTGCTACCTATGCGCTTCTTGCACATGTCCCTTC 493
 Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 Db 494 AAAGAAGGAACCTACTCACTGCCCAAGAGCAATTCGTTGTGCTGACCTGGAGCTGCC 553
 Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysArg 180
 Db 554 AGTTGGCTCACCACCGGGAACCTACCGATAGAGAGCTCTTGACAGCAGTGGGAAGCGT 613
 Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db 614 CTGGCTGCATCAAGATCGCTGCTCTCTAAAGGGCATA 652

RESULT 3
 BI820051
 LOCUS 603037236F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5178273 5',
 DEFINITION mRNA sequence.

ACCESSION BI820051
 VERSION BI820051.1 GI:15931601

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 858)
 NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cga@h-p@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLML1444 row: 9 column: 10

High quality sequence stop: 706.

Location/Qualifiers

1..858

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5178273"

/lab_host="DH10B"

/clone_lib="NIH_MGC_115"

/notes="Organ: pooled brain, lung, testis; Vector:

pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA

source anonymous pool of 6 male brains, age range 23-27; 1

male lung, age 27; and 1 male testis, age 69. Library is

oligo-dT primed and directionally cloned (EcoRV site is

destroyed upon cloning). Average insert size 1.8 kb,

insert size range 1-3 kb. Library is normalized and

enriched for full-length clones and was constructed by C.

Gruber (Invitrogen). Research Genetics tracking code

021. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 4.19e-91 Length: 858
 Score: 999.00 Matches: 190
 Percent Similarity: 98.45% Conservative: 0
 Best Local Similarity: 98.45% Mismatches: 3
 Query Match: 98.13% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BI820051 (1-858)

Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
 Db 8 ATGAGTCCCTGATGAGGCTCCCTCTGATCGCCCTGGGCTTGTCTCGCGTGCCT 67
 Qy 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 Db 68 GCGCAAGCCACCTGAAAGCCATCCAGCTCAGTAGCTTTCTCTGGGATACTGTGAT 127
 Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 Db 128 GAAGGGGAAGGACCCCTGCGGTGATCAGAAGGCTGACTCTGGAGCCTGACCCCATCGTGT 187
 Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 Db 188 CTTGGAATGTGACCTCAGTGTGCTGGGAGCACCAGTGTCCCTCTGAGTTCTCTCTG 247
 Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuThrPheLysIleProCysThr 100
 Db 248 AAGGTGGAATTTAGTTTGGAGAAGGAGTGGCTCTGGATCAAGATCCCATGCACA 307
 Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 Db 308 GACTACATTTGGCAGCTGTACCTTTGAACACTTCTGTGATGCTGTGACATGTTAATTCCT 367
 Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 Db 368 ACTGGGAGCCCTGCCAGAGCCCTCGTACCTATGGGCTTCTCTGCCACTGTCCCTTC 427
 Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
 Db 428 AAAGAAGGAACCTACTACCTGCCAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 487
 Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysArg 180
 Db 488 AGTTGGCTCACCACCGGGAACCTACCGCATAGAGAGCTCTGAGCAGCAGTGGGAAGCGT 547
 Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 Db 548 CTGGCTGCATCAAGATCGCTGCTCTCTAAAGGGCATA 586

RESULT 4

CR626644

LOCUS

DEFINITION

full-length cDNA clone CS0DL003VG14 of B cells (Ramos cell line)

Cot 25-normalized of Homo sapiens (human).

CR626644

ACCESSION

VERSION

CR626644.1 GI:50507451

KEYWORDS

HTC; CNSLT cDNA.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2338)

AUTHORS

Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished

REMARK

Contact: Feng Liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue

2 (bases 1 to 2338)

Genoscope.

Direct Submission

Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
Location/Qualifiers
source
1..2338
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL003YG14"
/tissue type="B cells (Ramos cell line) Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Alignment Scores:
Pred. No.: 3,74e-90 Length: 2338
Score: 996.00 Matches: 188
Percent Similarity: 98.45% Conservative: 2
Best Local Similarity: 97.41% Mismatches: 3
Query Match: 97.84% Indels: 0
DB: 3 Gaps: 0

US-10-030-937-9 (1-193) x CRG26644 (1-2338)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
DB 24 ATGCAGTCCCTGATGAGGCTCCCTCCTGATCGCCCTGGGCTTGCTTCGGGGCCCT 83

QY 21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
DB 84 GCGCAGGCCACCTGAGAAAGCATCCAGCTCAGTAGCTTTCTCTGGGATTAAGTGTGAT 143

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
DB 144 GAAGGGAAGGACCTCGGTGATCAGAGCTGACTCTGGAGGCTGACCCCATCATCGTT 203

QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
DB 204 CTGGGAATGTGACCTCAGTGTATGGGAGACACAGTGTCCCTCCTGATTCCTCTG 263

QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysLeuProCysThr 100
DB 264 AAGTGGATTAGTTTGGAGAGGAGGTGGCTCTGGATCAAGATCCCATGCACA 323

QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro 120
DB 324 GACTACATTGGCAGCTGTACCTTTGAACACTTGTGTGTGCTTGACATGTTAAATCCT 383

QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
DB 384 ACTGGGAGCCCTGCCAGAGCCCTCGGTACCTATGGGCTTCTTGCCACTGTCCCTTC 443

QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
DB 444 AAAGAAGGAACCTACTACTCTGCCCAAGAGCGAATTCGTTGTGCTGACCTGGAGCTGCC 503

QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysArg 180
DB 504 AGTTGGCTCACCACCGGAACTACCGCATAGAGGCTCTCTGAGCAGCAGTGGGAACGT 563

QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
DB 564 CTGGGCTGCATCAAGATCGTGTCTCTCTAAAGGCATA 602

RESULT 5
BU556606
LOCUS
DEFINITION
AGENCY 10188553 NIH MGC 109 Homo sapiens cDNA clone
IMAGE:6584324 5', mRNA sequence.
BU556606
ACCESSION

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BU556606.1 GI:22906878
EST.
Homo sapiens (human)
Homo sapiens
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 948)
NIH-MGC http://mgs.hci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2794 row: h column: 20
High quality sequence stop: 645.
Location/Qualifiers
source
1..948
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6584324"
/tissue type="teratocarcinoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone lib="NIH MGC 109"
/note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 1.57e-90 Length: 948
Score: 994.00 Matches: 189
Percent Similarity: 97.93% Conservative: 0
Best Local Similarity: 97.93% Mismatches: 4
Query Match: 97.64% Indels: 0
DB: 5 Gaps: 0

US-10-030-937-9 (1-193) x BU556606 (1-948)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
DB 23 ATGCAGTCCCTGATGAGGCTCCCTCCTGATCGCCCTGGGCTTGCTTCGGGGCCCT 82

QY 21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
DB 83 GCGCAGGCCACCTGAGAAAGCATCCAGCTCAGTAGCTTTCTCTGGGATTAAGTGTGAT 142

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
DB 143 GAAGGGAAGGACCTCGGTGATCAGAGCTGACTCTGGAGGCTGACCCCATCATCGTT 202

QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
DB 203 CTGGGAATGTGACCTCAGTGTCTGGGAGGAGGTGGCTCTGGATCAAGATCCCATGCACA 322

QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysLeuProCysThr 100
DB 263 AAGTGGATTAGTTTGGAGAGGAGGTGGCTCTGGATCAAGATCCCATGCACA 322

QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro 120
DB 323 GACTACATTGGCAGCTGTACTTTGAACACTTCTGTGTGCTGTGATGCTTAAATTCCT 382

```

QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 383 ACTGGGAGCCCTGCCAGAGCCCTGCGTACCTATGGCTTCTTGGCCACTGTCCTTC 442
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
Db 443 AAAGAGGAACCTACTCACTGCCCAAGAGCGAATTCGTTGTGCCGACCTGGAGCTGCC 502
QY 161 SerTrpLeuThrThrGlyAsnTyrArgileGluSerValLeuSerSerSerGlyLysArg 180
Db 503 AGTTGGCTCACACCGGGAATACCGCATAGAGCGCTCTGAGCAGCAGTGGGAAGCGT 562
QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 563 CTGGCTGCATCAAGATCGCTGCTCTCTANAGGCATA 601

RESULT 6
BQ643369
LOCUS AGENCOURT_8485451 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:6305296
DEFINITION 5', mRNA sequence.
ACCESSION BQ643369
VERSION BQ643369.1 GI:21767541
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 909)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2527 row: f column: 17
High quality sequence stop: 591.
FEATURES
source
Location/Qualifiers
1..909
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6305296"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_99"
/notes="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN
Alignment Scores:
Pred. No.: 4,77e-90 Length: 909
Score: 989.00 Matches: 188
Percent Similarity: 97.41% Conservative: 0
Best Local Similarity: 97.41% Mismatches: 5
Query Match: 97.15% Indels: 0
DB: 5 Gaps: 0

US-10-030-937-9 (1-193) x BQ643369 (1-909)

```

```

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuAlaThrPro 20
Db 6 ATGCAGTCCCTGATGAGGCTCCCTCTCTGATCGCCCTGGGCTTGCTTCTCGGGCCCT 65
QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTyrAspAsnCysPhe 40
Db 66 GCGCAAGCCCACTCGAAAGAGCCATCCAGCTCCAGGAGCTTTCTCTGGGATAACTGTGAT 125
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 126 GAAGGGAGAGACCTGCGGAGTCAAGAGCCCTGACTCTGGAGCCCTGACCCCATCGTCGT 185
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 186 CCTGGAAATGTGACCCCTCAGTGTGCTGGGAGCAGCAGTGTCCCTCAGTTCCTCTCTG 245
QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
Db 246 AAGGTGATTTAGTTTGGAGAGAGGAGTGGCTTGGCTTCTGGATCAAGATCCCATGCACA 305
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
Db 306 GACTACATTTGGCAGCTGTACCTTTGAACACTTCTGTGATGCTGTGACATGTTAATTCCT 365
QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
Db 366 ACTGGGAGCCCTGCCAGAGCCCTGCTGCTACCTATGGCTTCTTGGCCACTGTCCTTC 425
QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
Db 426 AAAGAGGAACCTACTCACTGCCCAAGAGCGAATTCGTTGTGCCCTGACCTGGAGCTGCC 485
QY 161 SerTrpLeuThrThrGlyAsnTyrArgileGluSerValLeuSerSerSerGlyLysArg 180
Db 486 AGTTGGCTCACACCGGGAATACCGCATAGAGCGCTCTGAGCAGCAGTGGGAAGCGT 545
QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 546 CTGGCTGCATCAAGATCGCTGCTCTCTAAAGGCATA 584

RESULT 7
BQ474816
LOCUS AGENCOURT_6476614 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562559
DEFINITION 5', mRNA sequence.
ACCESSION BQ474816
VERSION BQ474816.1 GI:18523858
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 973)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2292 row: c column: 08
High quality sequence stop: 603.
FEATURES
source
Location/Qualifiers
1..973
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

```

```
/clone="IMAGE:5562559"
/tissue type="duodenal adenocarcinoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone lib="NIH_MGC 88"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 5,27e-90 Length: 973
Score: 989.00 Matches: 188
Percent Similarity: 97.92% Conservativeness: 0
Best Local Similarity: 97.92% Mismatches: 4
Query Match: 97.15% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BM474816 (1-973)
QY 2 GlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrProAla 21
Db 1 CAGTCCTGATGAGGCTCCCTCTGATCGGCTGGGCTTCTCGGGCCCTGG 60
QY 22 GlnAlaHisLeuLysProSerGlnLeuSerSerPheSerTrpAspPheGlu 41
Db 61 CAAAGCCACCTGAAAGGATCCAGCTCAGTAGCTTTCTGGGATAACTGTGATGA 120
QY 42 GlyLysAspProAlaValLeuArgSerLeuThrLeuGluProAspProLeuValPro 61
Db 121 GGGAGGACCTCGCGGTGATCAGAGGCTGACTCTGGAGCCTGACCCATCGGTCT 180
QY 62 GlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeuLys 81
Db 181 GGAATGACCTCAGTGTCTGGGAGCAGCAGTGTCCCTGAGTCTCTCTCTGAG 240
QY 82 ValAspLeuValLeuGlyLysGluValAlaGlyLeuThrLysLysLeuProCysThrAsp 101
Db 241 GTGGATTAGTTTGGAGAGGAGGTGGCTGGCTCTGGATCAAGATCCCATGCAGAC 300
QY 102 TyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuProThr 121
Db 301 TACATTGGCAGCTGTACCTTTGAACACTCTGTGTGTGTGATGCTTGCATGTTAATCTACT 360
QY 122 GlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPheLys 141
Db 361 GGGAGCCCTGCGCAGAGCCCTCGTACTATGGGCTTCTTGGCCACTGTCCCTCAA 420
QY 142 GluGlyThrTyrSerLeuProLysSerSerGluPheAlaValProAspLeuGluLeuProSer 161
Db 421 GAAGGAACCTACTCACTGCCCAAGAGCGAATTGTTGTGCTGACCTGGAGCTGCCAGT 480
QY 162 TrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArgLeu 181
Db 481 TGCTCACCACCGGGAACTACCGCATAGAGGCTCTCTGAGCAGCTGGGAGGCTCTG 540
QY 182 GlyCysIleLysIleAlaAlaSerLeuLysGlyLeu 193
Db 541 GGCTGCATCAAGATCGCTGCTCTCTANAGGGCATA 576

RESULT 8
BM723945
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BM723945
UI-E-E01-aix-1-18-0-UI.r1 UI-E-E01 Homo sapiens cDNA clone
UI-E-E01-aix-1-18-0-UI 5', mRNA sequence.
BM723945
BM723945.1 GI:19045276
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 651)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
PUBMED
79044477
889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.
Location/Qualifiers
1. .651
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-E01-aix-1-18-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-E01"
/note="Organ: eye; Vector: p7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-E01 is a normalized cDNA library containing the
following tissue(s): fetal eye. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into p7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(3')18 tail. The sequence tag for this library is
CCGCTATACC. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI)."
```


184 TCCTGAAATGTGACCTCAGTGTGCTGGGAGCAGCAGTGTCCCTCCTCT 243
 80 uLyValAspLeuValLeuGluValAlaGlyLeuTrpLeuLysLeuProCysTh 100
 244 GAAGGTGGATTAGTTTGGAGAGAGGAGTGGCTGGCTTGGATCAAGATCCCATGCAC 303
 100 rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPr 120
 304 AGACTACATTGGCAGCTGACTTGAACACTTCTGTGATGTGCTTGACATGTAATTC 363
 120 oThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPh 140
 364 TACTGGGGAGCCTGCCAGAGCCCTGCGTACCTATGGCTTCTTGGCACTGTCCCTT 423
 140 eLyGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPr 160
 424 CAAGAAGGAACCTACTACTGCCCAAGAGCGAAATTCGTTGTGCTGACCTGGAGCTGCC 483
 160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysAr 180
 484 CAGTTGGCTACCAACCGGGAACCTACCCATAGAGAGCGTCTGTGAGCAGCTGGGAAGCG 543
 180 gLeuGlyCysIleLysLeuAlaAlaSerLeuLysGlyLe 193
 544 TCTGGGCTGCATCAAGATCGTGCCTCTCTAAAGGGCATA 583

RESULT 9

LOCUS BG762599 726 bp mRNA linear EST 15-MAY-2001
 DEFINITION 602734472F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4860136 5',
 mRNA sequence.

ACCESSION

VERSION BG762599.1 GI:14073252

KEYWORDS

SOURCE EST.

ORGANISM

Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 726)
 NIH-MGC <http://mgi.nci.nih.gov/>.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLCM1716 row: o column: 17

High quality sequence stop: 724.

FEATURES

source

1..726
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4860136"
 /tissue_type="melanotic melanoma, high MDR (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 49"
 /note="Organ: skin; Vector: pOT7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library. |"

ORIGIN

Alignment Scores:
 Pred. No.: 1.4e-89 Length: 726
 Score: 983.00 Matches: 185
 Percent Similarity: 98.43% Conservative: 3
 Best Local Similarity: 96.86% Mismatches: 3
 Query Match: 96.56% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BG762599 (1-726)

QY 3 SerLeuMetGlnAlaProLeuLeuAlaLeuGlyLeuLeuAlaThrProAlaGln 22
 Db 1 GCCCTGATGAGGCTCCCTCTCTGATCGCCCTGGGCTGTCTTCGGGGCCCTCGCGCA 60
 QY 23 AlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPheGluGly 42
 Db 61 GCCCACCTGAAAGCCATCCAGCTCAGTAGCTTTCTGGGATTAATGTGATGAAGGG 120
 QY 43 LysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValValProGly 62
 Db 121 AAGACCCCTCGGTGATCAGAAGCCTGACTCTGGAGCTGACCCCATCATCGTTCTCTGA 180
 QY 63 AsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeuLysVal 82
 Db 181 AATGTGACCTCAGTGTCTATGGGCGAGCAGGTGTCCCTCTGAGTTCTCTCTGAAGTG 240
 QY 83 AspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysLeuProCysThrAspTyr 102
 Db 241 GATTAGTTTGGAGAGGAGTGGCTGGGCTCTGGATCAAGATCCATGCACAGACTAC 300
 QY 103 IleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuProThrGly 122
 Db 301 ATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTTGACATGTTAATTCCTACTGG 360
 QY 123 GluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPheLysGlu 142
 Db 361 GAGCCCTGCCAGAGCCCTGCGTACTATAGGCTTCTCTGGCTGCTCCCTCTCAAGAA 420
 QY 143 GlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuProSerTrp 162
 Db 421 GGAACCTACTACTGCCAAGAGAGGAATTCGTGTGCTGCTGAGCTGCCAGCTGGCCAGTTGG 480
 QY 163 LeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArgLeuGly 182
 Db 481 CTCACCCCGGGAACCTACCGCATAGAGAGCTCTCGAGCAGCAGTGGGAAGCGTCTGGGC 540
 QY 183 CysIleLysIleAlaAlaSerLeuLysGlyLe 193
 Db 541 TGCATCAAGATCGTGCCTCTCTAAAGGGCATA 573

RESULT 10

LOCUS

BI838554

603086219F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5225092 5',
 mRNA sequence.

DEFINITION

BI838554

ACCESSION

VERSION BI838554.1 GI:15950104

KEYWORDS

SOURCE EST.

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 784)

NIH-MGC <http://mgi.nci.nih.gov/>.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov plate: LLAM11566 row: f column: 05 High quality sequence stop: 772.		found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov plate: LLAM11566 row: f column: 05 High quality sequence stop: 772.	
FEATURES	source	FEATURES	source
	1. .784 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5225092" /lab_host="DH10B" /clone_lib="NIH_MGC 120" /note="Organ: pooled pancreas and spleen; Vector: pCMV-SPT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH_MGC Library."		1. .949 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5816227" /tissue_type="lymphoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC 99" /note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN		ORIGIN	
Alignment Scores: Pred. No.: 1.98e-89 Length: 784 Score: 982.00 Matches: 189 Percent Similarity: 97.94% Conservative: 1 Best Local Similarity: 97.42% Mismatches: 3 Query Match: 96.46% Indels: 1 DB: 4 Gaps: 0		Alignment Scores: Pred. No.: 2.62e-89 Length: 949 Score: 982.00 Matches: 186 Percent Similarity: 98.41% Conservative: 0 Best Local Similarity: 98.41% Mismatches: 3 Query Match: 96.46% Indels: 0 DB: 5 Gaps: 0	
US-10-030-937-9 (1-193) x BI838554 (1-784)		US-10-030-937-9 (1-193) x BQ060062 (1-949)	
QY	1 MetGlnSerLeuMetGlnAlaProLeuLeuAlaLeuGly-LeuLeuLeuAlaThrPr 20	QY	5 MetGlnAlaProLeuLeuAlaLeuGlyLeuLeuAlaThrProAlaGlnAlaHis 24
Db	4 ATGCAGTCCCTGATGAGGCTCCCTCTGATCGCCCTGGGCATGCTTCTCGCGGCC 63	Db	3 ATGCAGCTCCCTCTCGATCGCCCTGGGCTTCTCGCGGCCCTCGCGAGCCAC 62
QY	20 oAlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPh 40	QY	25 LeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPheGluGlyLysAsp 44
Db	64 TGGCAAGCCACCTCAGAGAGCCATCCAGCTCAGTAGCTTTTCTCGGATAAATGTGA 123	Db	63 CTGAAAGACCATCCAGCTCAGTAGCTTTTCTGGGATAAATGTGTAAGGAGGAC 122
QY	40 eGluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVa 60	QY	45 ProAlaValIleArgSerLeuThrLeuGluProAspProIleValValProGlyAsnVal 64
Db	124 TGAAGGAGAGGACCTCGCGGTGATCAGAGCCCTGACTCTGGAGCCTGACCCCATCTCGT 193	Db	123 CCTGGGTGATCAGAGCCTGACTCTGGAGCCTGACCCCATCTCGTCTCTGGAATGTG 182
QY	60 lProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProle 80	QY	65 ThrLeuSerValValGlySerThrSerValProLeuSerSerProLeuLysValAspLeu 84
Db	184 TCCTGGAAATGTGACCTCAGTGTCTGGGACGACAGGTGCCCTGAGTTCTCTCT 243	Db	183 ACCCTCAGTGTCTGGGACGACAGGTGCCCTGAGTTCTCTCTGAGGTGATTTA 242
QY	80 uLysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysTh 100		
Db	244 GAAAGTGGATTAGTTTGGAGAAGAGGAGTGGCTCTCGATCAAGATCCCATGAC 303		
QY	100 rAspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePr 120		
Db	304 AGACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGCTGTGACATGTAATCC 363		
QY	120 oThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProph 140		
Db	364 TACTGGGAGCCCTGCCAGAGCCCTCGTACCTATGGGCTTCTTGGCACTGTCCCTT 423		
QY	140 eLysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPr 160		
Db	424 CAAAGAGAGAACCTACTACTCGCCCAAGAGCAATTCGTTGTGCTGACCTGGAGCTGCC 483		
QY	160 oSerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysAr 180		
Db	484 CAGTTGGCTCACCCACCGGAACTACCGCATAGAGAGGCTCCTGAGCAGCAGTGGAGG 543		
QY	180 gLeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193		
Db	544 TCTGGGCTGCATCAAGATCGTGCCTCTCTTAAGAGGCATA 583		

QY 85 ValLeuGluLysGluValAlaGlyLeuTrpIleLysLeuProCysThrAspTyrIleGly 104
 Db 243 GTTTTGGAGAGAGGTGGCTGCTCTGGATCAGATCCCATGACACTACATTTGGC 302
 QY 105 SerCysThrPheGluHisPheCysAspValLeuAspMetLeuIleProThrGlyGluPro 124
 Db 303 AGCTGTACCTTTGAACACTTCTGTGTGCTGATGCTGATGCTTAATTCCTACTGGGAGGCC 362
 QY 125 CysProGluProLeuAlaGlyThrTyrGlyLeuProCysHisCysProPheLysGluGlyThr 144
 Db 363 TGCCACAGAGCCCTCGGTACCTATGCGGTTCCTTGCCACTGTCCTTCAAAGAGGAAC 422
 QY 145 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuProSerTrpLeuThr 164
 Db 423 TACTCACTGCCCAAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCCAGTTGGCTCACC 482
 QY 165 ThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysArgLeuGlyCysIle 184
 Db 483 ACCGGGAACCTACCGCATAGAGCGTCTGAGCAGCAGTGGGAGCGTCTGGGCTGCATC 542
 QY 185 LysIleAlaAlaSerLeuLysGlyIle 193
 Db 543 AAGATCGCTGCTCTCTAAAGGGGCATA 569

RESULT 12
 BG323734 784 bp mRNA linear EST 27-FEB-2001
 LOCUS 60242183371 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4559888 5',
 DEFINITION mRNA sequence.

ACCESSION
 VERSION BG323734.1 GI:13130171

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI268 row: a column: 09

High quality sequence stop: 777.

Location/Qualifiers

FEATURES

SOURCE

1. .784

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4559888"

/tissue_type="renal cell adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: kidney; Vector: pOTB7; Site:1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.: 2.5e-89

Score: 981.00

Length: 784

Matches: 191

Percent Similarity: 98.96% Conservative: 0
 Best Local Similarity: 98.96% Mismatches: 2
 Query Match: 96.37% Indels: 1
 DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BG323734 (1-784)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuAlaThrPro 20

Db 7 ATGCAGTCCCTGATCAGGCTCCCTCTCTGATCGCCCTGGGCTTCTTCGCGGC-CCT 65

QY 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspSerCysPhe 40

Db 66 GCGCAAGCCCACTGAAAAGCCATCCAGGCTCAGTAGCTTTCTCTGGGATAACTGTGAT 125

QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60

Db 126 GAAGGGAAGGACCTGCGGTGATCAGAAGCTGACTCTGGAGCCTGACCCCATCGTCGT 185

QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80

Db 186 CCTGGAATGTGACCCCTCAGTGTGCTGGGAGCACCAGTGTCCCTCTGAGTTCTCTCTG 245

QY 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100

Db 246 AAGTGGGATTTAGTTTGGAGAAGGAGGTGGCTGCGCTCTGGATCAAGATCCCATGACA 305

QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120

Db 306 GACTACATTTGCGAGCTGTACCTTTGAACACTTCTGTGAIGTGTGATGTTAATTCCT 365

QY 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140

Db 366 ACTGGGAGGCGCTGCCAGAGCCCTCGCTACCTATGGCTTCTTCCCTGACCTGTCCTTC 425

QY 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160

Db 426 AAAGAAGGAACCTACTCACTGCCCAAGAGCGAATTCGCTGTGCTGACCTGGAGTGCCC 485

QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180

Db 486 AGTTGGTCTACCCAGGGAACCTACCGCATAGAGCGTCTTGAGCAGCGTGGGAAGCGT 545

QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193

Db 546 CTGGGCTGCATCAAGATCGCTGCTCTCTCTAAAGGGGCATA 584

RESULT 13

BI856212

LOCUS 603382936F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5391801 5',

DEFINITION mRNA sequence.

ACCESSION BI856212

VERSION BI856212.1 GI:15996959

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 733)

NIH-MGC http://imgc.ncbi.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM11998 row: h column: 10

High quality sequence start: 2
High quality sequence stop: 733.

FEATURES
source
1. .733
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5391801"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="NIH MGC 87"
/clone_lib="NIH MGC 87"
/note="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 5,78e-89 Length: 733
Score: 977.00 Matches: 190
Percent Similarity: 98.45% Conservatives: 0
Best Local Similarity: 98.45% Mismatches: 3
Query Match: 95.97% Indels: 1
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x B1856212 (1-733)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db 21 ATGCAGTCCCTGATGAGGCTCCCTCTGATCGCCCTGGGCTTGTCTTCGCGGC-CCT 79
QY 21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerThrPheAsnCysPhe 40
Db 80 GCGCAAGCCACCTGAAAGCCATCCAGTCAGTAGCTTTTCTCTGGGATACTGTGAT 139
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 140 GAAGGGAAGACCTCGCGGTGATCAGAAGCCTGACTCTGGAGCCTGACCCATCGTCT 199
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 200 CTGGGAATTTAGTCTTGGAGAGAGGCTGGCTGCTGATGAGTCTTCTCTCTG 259
QY 81 LysValAspLeuValLeuGlyLysGluValAlaGlyLeuThrPheLysIleProCysThr 100
Db 260 AAGGTGATTTAGTTTGGAGAGAGGCTGGCTGCTGATGAGTCTTCTCTCTG 319
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro 120
Db 320 GACTACATTTGGAGTCTTGTGATGAGTCTTGTGATGAGTCTTGTGATGAGTCTTGT 379
QY 121 ThrGlyGluProCysProGluProLeuArgThrThrGlyLeuProCysHisCysProPhe 140
Db 380 ACTGGGAGGCTCCAGAGCCCTGGTACCTATGCGCTTCTTCCACTGTCCCTTC 439
QY 141 LysGluGlyThrThrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 160
Db 440 AAAGAGGAACCTACTCACTGCCCAAGAGCAATTCGTTGCTGCTGACCTGGAGCTGCC 499
QY 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerGlyLysArg 180
Db 500 AGTGGCTTACACCGGGAATCTACCGATAGAGGCTGCTTCGACAGCAGTGGAGCGT 559
QY 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
Db 560 CTGGGCTGCATCAGATCGTCTCTCTAAAGGGCAATA 598

RESULT 14
BI091220 842 bp mRNA linear EST 20-JUN-2001
LOCUS 602856051F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4997365 5',
DEFINITION mRNA sequence.

ACCESSION BI091220
VERSION BI091220.1 GI:14509550
EST.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1024 row: e column: 14
High quality sequence stop: 739.
Location/Qualifiers
1. .842
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4997365"
/cell_line="MGC36"
/lab_host="NIH MGC 10"
/clone_lib="NIH MGC 10"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5 kb. Library prepared by Life Technologies."

ORIGIN
Alignment Scores:
Pred. No.: 7,09e-89 Length: 842
Score: 977.00 Matches: 189
Percent Similarity: 97.93% Conservatives: 0
Best Local Similarity: 97.93% Mismatches: 4
Query Match: 95.97% Indels: 1
DB: 4 Gaps: 0

US-10-030-937-9 (1-193) x BI091220 (1-842)

QY 1 MetGlnSerLeuMetGlnAlaProLeuLeuLeuAlaLeuGlyLeuLeuAlaThrPro 20
Db 12 ATGCAGTCCCTGATGAGGCTCCCTCTGATCGCCCTGGGCTTGTCTTCGCGGCCCT 71
QY 21 AlaGlnAlaHisLeuLysProSerGlnLeuSerSerPheSerThrPheAsnCysPhe 40
Db 72 GCGCAAGCCACCTGAAAGCCATCCAGTCAGTAGCTTTTCTCTGGGATACTGTGAT 131
QY 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
Db 132 GAAGGGAAGACCTCGCGGTGATCAGAAGCCTGACTCTGGAGCTGACCCATCGTCT 191
QY 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
Db 192 CTGGGAATTTAGTCTTGGAGAGAGGCTGGCTGCTGATGAGTCTTCTCTCTCTG 251
QY 81 LysValAspLeuValLeuGlyLysGluValAlaGlyLeuThrPheLysIleProCysThr 100
Db 252 AAGGTGATTTAGTTTGGAGAGAGGCTGGCTGCTGATGAGTCTTGTGATGAGTCTTGT 311
QY 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuLeuPro 120
Db 312 GACTACATTTGGAGTCTTGTGATGAGTCTTGTGATGAGTCTTGTGATGAGTCTTGT 371
QY 121 ThrGlyGluProCysProGluProLeuArgThrThrGlyLeuProCysHisCysProPhe 140

Db 372 ACTGGGGAGCCCTGCCAGAGCCCTCGTACCTATGCGCTTCTTGGCCACTGTCCCTTC 431

Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuLeuPro 160
 |||||
 Db 432 AAAGAGGAACTACTACTGCTCCAGAGCGAATTGCTTGTGCTCCTGACCTGGAGCTGCC 491

Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 |||||
 Db 492 AGTTGGCTCACCACCGGAACTACCGCATAGAGAGCGTCTGAGCAGCAGTGGGAGCGT 551

Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 |||||
 Db 552 CTGGGCTGCATCAAGATCGCTGC-TCTCTAAAGGGCATA 589

RESULT 15
 LOCUS AL543858
 DEFINITION AL543858 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
 clone CS0DI005YG11 5-PRIME, mRNA sequence.
 ACCESSION AL543858
 VERSION AL543858
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 912)
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On Feb 15, 2001 this sequence version replaced gi:31265703.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3172.f

For more information about this cluster, see
<http://www.genoscope.cns.fr/cdna?s=CS0DI005AD06QPI&c=3172.f>.

FEATURES

source
 1. .912
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DI005YG11"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 7.98e-89 Length: 912
 Score: 977.00 Matches: 189
 Percent Similarity: 98.45% Conservative: 1
 Best Local Similarity: 97.93% Mismatches: 3
 Query Match: 95.97% Indels: 1
 DB: 1 Gaps: 0

US-10-030-937-9 (1-193) x AL543858 (1-912)

Qy 1 MetGlnSerLeuMetGlnAlaProLeuLeuIleAlaLeuGlyLeuLeuAlaThrPro 20
 |||||
 Db 9 ATGCAGTCCCTGATGAGGCTCCCTCTGATCGCCCTGGGCTTGTCTCGGACCCCT 68
 |||||
 Qy 21 AlaGlnAlaHisLeuLysLysProSerGlnLeuSerSerPheSerTrpAspAsnCysPhe 40
 |||||

Db 69 GCGCAASC-CACCTGAAAAAGCCATCCAGCTCAGTAGCTTTTCTCTGGGATAACTGTGAT 127

Qy 41 GluGlyLysAspProAlaValIleArgSerLeuThrLeuGluProAspProIleValVal 60
 |||||
 Db 128 GAAGGGAAGGACCCCTGCGGTGATCAGAACCTGACTCTGGAGCCTGACCCCATCATCGTT 187

Qy 61 ProGlyAsnValThrLeuSerValValGlySerThrSerValProLeuSerSerProLeu 80
 |||||
 Db 188 COTGAAATGTGACCTCAGTGTCTGTGGGCAGCACAGTGTCCCTGAGTTCTCTCTCTG 247

Qy 81 LysValAspLeuValLeuGluLysGluValAlaGlyLeuTrpIleLysIleProCysThr 100
 |||||
 Db 248 AAGTGGATTAGTTTTGGAGAAGGAGGTGGCTGGCCTCTGGATCAAGATCCCATGCACA 307

Qy 101 AspTyrIleGlySerCysThrPheGluHisPheCysAspValLeuAspMetLeuIlePro 120
 |||||
 Db 308 GACTACATTGGCAGCTGTACCTTTGAACACTTCTGTGATGTGCTTGACATGTTAATTCCT 367

Qy 121 ThrGlyGluProCysProGluProLeuArgThrTyrGlyLeuProCysHisCysProPhe 140
 |||||
 Db 368 ACTGGGAGCCCTGCCAGAGCCCTGCGTACTATGGGCTTCTCTGGCCTGCTCCCTTC 427

Qy 141 LysGluGlyThrTyrSerLeuProLysSerGluPheAlaValProAspLeuLeuPro 160
 |||||
 Db 428 AAAGAGGAACCTACTCTGCTCCCAAGAGCGAATTCTGTGTGCTGACCTGGAGCTGCC 487

Qy 161 SerTrpLeuThrThrGlyAsnTyrArgIleGluSerValLeuSerSerSerGlyLysArg 180
 |||||
 Db 488 AGTTGGCTCACCACCGGAACTACCGCATAGAGAGCGTCTCTGAGCAGCAGTGGGAGCGT 547

Qy 181 LeuGlyCysIleLysIleAlaAlaSerLeuLysGlyIle 193
 |||||
 Db 548 CTGGGCTGCATCAAGATCGCTGCCTCTCTAAAGGGCATA 586

Search completed: November 19, 2004, 00:22:22
 Job time : 6216.45 secs

This Page Blank (usp10)

Result No.	Query %		DB	ID	Description
	Score	Match Length			
1	1018	100.0	193	4	AAB31897
2	1005	98.7	193	4	AAB31901
3	1005	98.7	193	4	AAB31902
4	1005	98.7	193	4	AAB31904
5	1005	98.7	193	4	AAB31896
6	1005	98.7	193	4	AAB31928
7	1005	98.7	193	4	AAB31903
8	1005	98.7	193	5	ABG31345
9	1005	98.7	193	5	ABG65212
10	1005	98.7	193	7	ADN95858
11	1005	98.7	193	8	ADN03620
12	1000	98.2	193	4	ABG00720
13	1000	98.2	193	8	ADQ17712
14	986	97.8	200	4	AAB31899
15	987	97.0	189	4	AAB31900
16	936	91.9	178	4	AAB31898
17	857	84.2	178	5	ABG31346
18	739.5	72.6	199	2	AAW10656
19	419	41.2	76	4	ABG00717
20	412	40.5	191	5	ABP41479
21	353	34.7	61	4	AAW15082
22	353	34.7	61	4	ABB34073
23	353	34.7	61	4	AAW27530
24	353	34.7	61	4	ABB28897
25	353	34.7	61	4	ABB19511

CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 100.0%; Score 1018; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 4.7e-103;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOSLMOAPLLIALGILLATPAQAHLKKPSQSFSDWNCDEGKDPVIRSLTLEPPPIV 60
DB 1 MOSLMOAPLLIALGILLATPAQAHLKKPSQSFSDWNCDEGKDPVIRSLTLEPPPIV 60
QY 61 PGNVTLVVGVSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVLDMLIP 120
DB 61 PGNVTLVVGVSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVLDMLIP 120
QY 121 TGEPCPEPLRTYGLPCHCPKCTYSLPKSEFAVPDLELPSPWLTGNYRIESVLSGSKR 180
DB 121 TGEPCPEPLRTYGLPCHCPKCTYSLPKSEFAVPDLELPSPWLTGNYRIESVLSGSKR 180
QY 181 LGCIKIAASLKGI 193
DB 181 LGCIKIAASLKGI 193
RESULT 2
AAB31901
ID AAB31901 standard; protein; 193 AA.
AC AAB31901;
DT 15-MAY-2001 (first entry)
XX
XX Amino acid sequence of a human protein.
DE Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
XX Homo sapiens.
OS
PN WO200105422-A2.
XX
XX 25-JAN-2001.
XX
XX 17-JUL-2000; 2000WO-FR002057.
XX
XX 15-JUL-1999; 99FR-00009372.
XX
XX (INMR) BIOMERIEUX STELHYS.
XX
XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
XX Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
XX Claim 1; Page 162-163; 209pp; French.
XX
XX The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a

CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 98.7%; Score 1005; DB 4; Length 193;
Best Local Similarity 99.0%; Pred. No. 1.3e-101;
Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MOSLMOAPLLIALGILLATPAQAHLKKPSQSFSDWNCDEGKDPVIRSLTLEPPPIV 60
DB 1 MOSLMOAPLLIALGILLATPAQAHLKKPSQSFSDWNCDEGKDPVIRSLTLEPPPIV 60
QY 61 PGNVTLVVGVSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVLDMLIP 120
DB 61 PGNVTLVVGVSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVLDMLIP 120
QY 121 TGEPCPEPLRTYGLPCHCPKCTYSLPKSEFAVPDLELPSPWLTGNYRIESVLSGSKR 180
DB 121 TGEPCPEPLRTYGLPCHCPKCTYSLPKSEFAVPDLELPSPWLTGNYRIESVLSGSKR 180
QY 181 LGCIKIAASLKGI 193
DB 181 LGCIKIAASLKGI 193
RESULT 3
AAB31902
ID AAB31902 standard; protein; 193 AA.
AC AAB31902;
DT 15-MAY-2001 (first entry)
XX
XX Amino acid sequence of a human protein.
DE Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
XX Homo sapiens.
OS
PN WO200105422-A2.
XX
XX 25-JAN-2001.
XX
XX 17-JUL-2000; 2000WO-FR002057.
XX
XX 15-JUL-1999; 99FR-00009372.
XX
XX (INMR) BIOMERIEUX STELHYS.
XX
XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
XX Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
XX Claim 1; Page 163; 209pp; French.
XX
XX The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses

CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 98.7%; Score 1005; DB 4; Length 193;
 Best Local Similarity 99.0%; Pred. No. 1.3e-101;
 Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQSILMQAPLIIALLGLLIIATPAQHLKKPSQLSFSFSDNCFEGKDPVIRSLTLEPDIIV 60
 DB 1 MQSILMQAPLIIALLGLLIIATPAQHLKKPSQLSFSFSDNCFEGKDPVIRSLTLEPDIIV 60
 QY 61 PGNVTLISVVGSTSVPLSSPLKVDLVEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
 DB 61 PGNVTLISVVGSTSVPLSSPLKVDLVEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
 DB 121 TGEPCPEPLRTYGLPCHCPKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
 QY 181 LGCITKIAASLKI 193
 DB 181 LGCITKIAASLKI 193

RESULT 4
 AAB31904
 ID AAB31904 standard; protein; 193 AA.
 XX
 AC AAB31904;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of a human protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.
 XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX

PS Claim 1; Page 164-165; 209pp; French.
 XX
 CC The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 98.7%; Score 1005; DB 4; Length 193;
 Best Local Similarity 99.0%; Pred. No. 1.3e-101;
 Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQSILMQAPLIIALLGLLIIATPAQHLKKPSQLSFSFSDNCFEGKDPVIRSLTLEPDIIV 60
 DB 1 MQSILMQAPLIIALLGLLIIATPAQHLKKPSQLSFSFSDNCFEGKDPVIRSLTLEPDIIV 60
 QY 61 PGNVTLISVVGSTSVPLSSPLKVDLVEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
 DB 61 PGNVTLISVVGSTSVPLSSPLKVDLVEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
 DB 121 TGEPCPEPLRTYGLPCHCPKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
 QY 181 LGCITKIAASLKI 193
 DB 181 LGCITKIAASLKI 193

RESULT 5
 AAB31896
 ID AAB31896 standard; protein; 193 AA.
 XX
 AC AAB31896;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of the human ganglioside GM2 activator protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.
 DR N-PSDB; AAF54698.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 23; Page 158-159; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 98.7%; Score 1005; DB 4; Length 193;
 Best Local Similarity 99.0%; Pred. No. 1.3e-101;
 Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MOSLQAPLLIALGLLLATPAQAHLKPKSOLFSSWDCNCFEGKDPVIRSLTEPPIV 60
 DB 1 MOSLQAPLLIALGLLLATPAQAHLKPKSOLFSSWDCNCFEGKDPVIRSLTEPPIV 60
 QY 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFFHFCVDLMLIP 120
 DB 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFFHFCVDLMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFAVPDLELPMSLTGNYRIESVLSGSKR 180
 DB 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFAVPDLELPMSLTGNYRIESVLSGSKR 180
 QY 181 LGCIXIAASLKG 193
 DB 181 LGCIXIAASLKG 193

RESULT 6
 AAB31928
 ID AAB31928 standard; protein; 193 AA.
 XX
 AC AAB31928;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of the human ganglioside GM2 activator protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99PR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.

XX
 PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 DR
 XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Disclosure; Fig 1; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 98.7%; Score 1005; DB 4; Length 193;
 Best Local Similarity 99.0%; Pred. No. 1.3e-101;
 Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MOSLQAPLLIALGLLLATPAQAHLKPKSOLFSSWDCNCFEGKDPVIRSLTEPPIV 60
 DB 1 MOSLQAPLLIALGLLLATPAQAHLKPKSOLFSSWDCNCFEGKDPVIRSLTEPPIV 60
 QY 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFFHFCVDLMLIP 120
 DB 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFFHFCVDLMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFAVPDLELPMSLTGNYRIESVLSGSKR 180
 DB 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFAVPDLELPMSLTGNYRIESVLSGSKR 180
 QY 181 LGCIXIAASLKG 193
 DB 181 LGCIXIAASLKG 193

RESULT 7
 AAB31903
 ID AAB31903 standard; protein; 193 AA.
 XX
 AC AAB31903;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of a human protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99FR-00009372.
 XX PA (INNR) BIOMERIEUX STELHYS.
 XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX DR WPI; 2001-159475/16.
 XX PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX PS Claim 1; Page 164; 209pp; French.
 XX CC The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX SQ Sequence 193 AA;
 Query Match 98.7%; Score 1005; DB 4; Length 193;
 Best Local Similarity 99.0%; Pred. No. 1.3e-101;
 Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIALGLLLATPAQAHKKPSQLSFSFMDNCFEGKDPVAVIRSLTLEPPIV 60
 Db 1 MQSLMQAPLLIALGLLLATPAQAHKKPSQLSFSFMDNCFEGKDPVAVIRSLTLEPPIV 60
 QY 61 PGNVTLVSVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
 Db 61 PGNVTLVSVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPKPGTYSLPKSEFAVPDLPLSWLTGNYRIESVLSSSGKR 180
 Db 121 TGEPCPEPLRTYGLPCHCPKPGTYSLPKSEFAVPDLPLSWLTGNYRIESVLSSSGKR 180
 QY 181 LGCIKIAASLKG 193
 Db 181 LGCIKIAASLKG 193
 RESULT 8
 ABG31345
 ID ABG31345 standard; protein; 193 AA.
 XX AC ABG31345;
 XX DT 15-NOV-2002 (first entry)
 XX DE Human GM2 activator protein.
 KW Human; GM2 activator protein; ganglioside; platelet activating factor;
 KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
 KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
 KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
 KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
 KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
 KW antiinflammatory; antiasthmatic; antiarthritic; antirheumatic;
 KW antiallergic; hepatotropic; nephrotropic; immunosuppressive;
 KW tranquiliser.

XX OS Homo sapiens.
 XX PN US6423680-B1.
 XX PD 23-JUL-2002.
 XX PF 30-OCT-1998; 98US-00183841.
 XX PR 30-OCT-1998; 98US-00183841.
 XX PA (HSCR-) HSC RES & DEV LP.
 XX PI Rigat B, Reynaud D, Mahuran D;
 XX DR WPI; 2002-664636/71.
 XX PT Composition useful for treating inflammatory conditions e.g. asthma
 PT comprises GM2 activator protein or GM2 activator peptide in combination
 PT with a carrier.
 XX CC Claim 3; Fig 1; l1pp; English.
 XX CC The present invention relates to a composition comprising GM2
 CC (ganglioside) activator protein or a GM2 activator peptide derived from
 CC the GM2 activator protein in combination with a carrier. The composition
 CC comprises the protein or peptide in an amount of 1-100 mg. The GM2
 CC activator protein is capable of inhibiting platelet activating factor
 CC (PAF). The composition of the invention is useful for treating
 CC inflammatory disorders e.g. inflammatory bowel disease, asthma,
 CC autoimmune disease (such as lupus), hypersensitivity infection,
 CC rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
 CC gout and tissue-specific conditions (e.g. glomerulonephritis and
 CC hepatitis). The composition is capable of inhibiting platelet activating
 CC factor, is non-toxic, is efficacious and presents less severe side
 CC effects, including redness, swelling, pain and polymorphonuclear
 CC leukocyte accumulation at the inflammatory site and other associated
 CC cellular responses. The present sequence represents human GM2 activator
 CC protein
 XX SQ Sequence 193 AA;
 Query Match 98.7%; Score 1005; DB 5; Length 193;
 Best Local Similarity 99.0%; Pred. No. 1.3e-101;
 Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MQSLMQAPLLIALGLLLATPAQAHKKPSQLSFSFMDNCFEGKDPVAVIRSLTLEPPIV 60
 Db 1 MQSLMQAPLLIALGLLLATPAQAHKKPSQLSFSFMDNCFEGKDPVAVIRSLTLEPPIV 60
 QY 61 PGNVTLVSVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
 Db 61 PGNVTLVSVGSTVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCFHFCDVLDMLIP 120
 QY 121 TGEPCPEPLRTYGLPCHCPKPGTYSLPKSEFAVPDLPLSWLTGNYRIESVLSSSGKR 180
 Db 121 TGEPCPEPLRTYGLPCHCPKPGTYSLPKSEFAVPDLPLSWLTGNYRIESVLSSSGKR 180
 QY 181 LGCIKIAASLKG 193
 Db 181 LGCIKIAASLKG 193
 RESULT 9
 ABP65212
 ID ABP65212 standard; protein; 193 AA.
 XX AC ABP65212;
 XX DT 12-NOV-2002 (first entry)
 XX DE Hypoxia-regulated protein #86.

KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;
 KW antiinflammatory; vulnery; gynecological; ophthalmological; vaccine;
 KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
 KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;
 KW preclampsia; atherosclerosis; inflammatory condition; wound healing;
 KW inflammation; erythropoiesis; hair loss; human.
 XX

OS Homo sapiens.
 XX WO200246465-A2.
 XX 13-JUN-2002.
 XX 10-DEC-2001; 2001WO-GB005458.
 XX 08-DEC-2000; 2000GB-00030076.
 PR 08-FEB-2001; 2001GB-00003156.
 PR 25-OCT-2001; 2001GB-00025666.
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.
 PA White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA,
 PI Rayner WN;
 PI WPI; 2002-627238/67.
 DR

XX Identifying a gene involved in disease for treating hypoxia-regulated
 XX conditions, comprises comparing the transcriptome/proteome of two cell
 XX types under different conditions and identifying a differentially
 XX regulated gene.
 XX Claim 35; Page 397; 538pp; English.
 XX

CC The present invention relates to methods for identifying genes and
 CC proteins that are implicated in a specific disease or physiological
 CC condition. The method comprises comparing the transcriptome/proteome of a
 CC specialised cell type implicated in a disease or condition with that of a
 CC second specialised cell type, under two experimental conditions, and
 CC identifying a gene that is differentially regulated in the two
 CC specialised cell types under experimental conditions. ABV77873-ABV78116
 CC and ABP65061-ABP65257 were identified using the methods of the invention.
 CC The coding sequences and proteins are useful for treating a disease in a
 CC patient, for manufacture of a medicament for treating hypoxia-regulated
 CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
 CC biological response to hypoxia conditions, or hypoxic-associated
 CC pathology in a patient. The coding sequences and proteins are also useful
 CC for monitoring the therapeutic treatment of a disease or physiological
 CC condition, such as cancer, ischaemic conditions, reperfusion injury,
 CC retinopathy, neonatal stress, preclampsia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss
 CC

SQ Sequence 193 AA;
 Query Match 98.7%; Score 1005; DB 5; Length 193;
 Best Local Similarity 99.0%; Pred. No. 1.3e-101;
 Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MQSLMQAPLLIALGILLATPAQAHKKPSQSSSFWDCNCFEGKDPVIRSLTEPDPV 60
 Db 1 MQSLMQAPLLIALGILLATPAQAHKKPSQSSSFWDCNCFEGKDPVIRSLTEPDPV 60
 Qy 61 PGNVTLVVGSTSVPLSSPLKVLVLEKEVAGLWIKIPCTDIYIGCTFEHFCVDLMLIP 120
 Db 61 PGNVTLVVGSTSVPLSSPLKVLVLEKEVAGLWIKIPCTDIYIGCTFEHFCVDLMLIP 120
 Qy 121 TGEPCPEPLRYGHPCHCPFFKGYSLPKSEFAPVDPDLPSWLTGNVRIESVLSGGR 180
 Db 121 TGEPCPEPLRYGHPCHCPFFKGYSLPKSEFAPVDPDLPSWLTGNVRIESVLSGGR 180
 Qy 181 LGCIKIAASLKI 193
 Db 181 LGCIKIAASLKI 193

RESULT 10
 ADN955858
 ID ADN955858 standard; protein; 193 AA.
 XX AC ADN955858;
 XX DT 01-JUL-2004 (first entry)
 XX DE Human BEC/LEC-related protein sequence SeqID782.
 XX KW growth; differentiation; blood endothelial cell; BEC;
 KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
 KW lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;
 KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
 KW inflammatory disease; cancer metastasis; lymphatic system; human.
 XX OS Homo sapiens.
 XX WO2003080640-A1.
 XX 02-OCT-2003.
 XX 07-MAR-2003; 2003WO-US006900.
 XX 07-MAR-2002; 2002US-0363019P.
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX (LICN) LICENTIA LTD.
 XX Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
 XX WPI; 2003-876899/81.
 XX N-PSDB; ADN955859.
 XX Example 1; SEQ ID NO 782; 176pp; English.

CC This invention relates to a method of differentially modulating the
 CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
 CC endothelial cells (LEC) comprises contacting endothelial cells with a
 CC composition comprising an agent that differentially modulates blood or
 CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
 CC identifying a human subject with lymphoedema and with a mutation in at
 CC least one allele of a gene encoding a LEC protein, where the mutation
 CC correlates with lymphoedema in human subjects, and with the proviso that
 CC the LEC protein is not VEGFR-3; and administering to the subject a
 CC composition comprising a lymphatic growth agent selected from VEGF-C or
 CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
 CC the development of compounds with an antiangiogenic, cytostatic,
 CC vasotropic or antiinflammatory activity or for gene therapy. The method
 CC is useful in modulating the growth or differentiation of blood
 CC endothelial cells or lymphatic endothelial cells, in treating hereditary
 CC lymphoedema, in screening for an endothelial cell disorder or
 CC a drug on endothelial cells. The agent is useful in manufacturing a
 CC medicament for the differential modulation of blood vessel endothelial
 CC cell or lymphatic vessel endothelial cell growth or differentiation. The
 CC lymphatic growth agent may also be used in manufacturing a medicament for
 CC the treatment of hereditary lymphoedema resulting from a mutation in a
 CC LEC gene or of other diseases involving the lymphatic vessels, such as
 CC various inflammatory diseases and cancer metastasis via the lymphatic
 CC system. The present sequence is that of a human LEC/BEC differentially
 CC expressed protein which is related to the method of the invention. Note:
 CC This sequence does not appear in the specification but was obtained by
 CC the indexer using the source data given in table 14 of the specification.

SQ Sequence 193 AA;

Query Match 98.7%; Score 1005; DB 7; Length 193;
 Best Local Similarity 99.0%; Pred. No. 1.3e-101;
 Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MQSLMQAPLLIALGILLATPAQAHKKPSQSSSFWDCNCFEGKDPVIRSLTEPDPV 60

Db 1 MQSLMQAPLLIALGULLATPAQAHKKPSQLSSFSWDCNDEGKDPVIRSLTLEPPIV 60
QY 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLIP 120
Db 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLIP 120
QY 121 TGEPCEPLRTYGLPCHCPKFGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
Db 121 TGEPCEPLRTYGLPCHCPKFGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
QY 181 LGCITKAASLKI 193
Db 181 LGCITKAASLKI 193
RESULT 11
ADN03620
ID ADN03620 standard; protein; 193 AA.
XX AC
ADN03620;
XX
01-JUL-2004 (first entry)
XX
Antipsoriatic protein sequence #7.
XX
antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
Homo sapiens.
XX
W02004028479-A2.
XX
08-APR-2004.
XX
25-SEP-2003; 2003WO-US030907.
XX
25-SEP-2002; 2002US-0414006P.
XX
(GETH) GENENTECH INC.
XX
Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
XX
WPI; 2004-305105/28.
DR N-PSDB; ADN03619.
XX
New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
XX
Claim 9; SEQ ID NO 14; 3069pp; English.
XX
The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polypeptides of the invention.
XX
Sequence 193 AA;
SQ
Query Match 98.7%; Score 1005; DB 8; Length 193;
Best Local Similarity 99.0%; Pred. No. 1.3e-101;
Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MOSLMOAPLLIALGULLATPAQAHKKPSQLSSFSWDCNDEGKDPVIRSLTLEPPIV 60
Db 1 MOSLMOAPLLIALGULLATPAQAHKKPSQLSSFSWDCNDEGKDPVIRSLTLEPPIV 60
QY 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLIP 120
Db 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLIP 120
QY 121 TGEPCEPLRTYGLPCHCPKFGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180

Db 121 TGEPCEPLRTYGLPCHCPKFGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
QY 181 LGCITKAASLKI 193
Db 181 LGCITKAASLKI 193
RESULT 12
ABG00720
ID ABG00720 standard; protein; 193 AA.
XX AC
ABG00720;
XX
13-FEB-2002 (first entry)
XX
Novel human diagnostic protein #711.
XX
Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
Homo sapiens.
XX
W0200175067-A2.
XX
11-OCT-2001.
XX
30-MAR-2001; 2001WO-US008631.
XX
31-MAR-2000; 2000US-00540217.
XX
23-AUG-2000; 2000US-00649167.
XX
(HYSE-) HYSEQ INC.
XX
Drmanac RT, Liu C, Tang YT;
PI
XX
WPI; 2001-639362/73.
DR N-PSDB; AAS64907.
XX
New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
Claim 20; SEQ ID NO 31079; 103pp; English.
XX
The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG030377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
Sequence 193 AA;
SQ
Query Match 98.2%; Score 1000; DB 4; Length 193;
Best Local Similarity 98.4%; Pred. No. 4.4e-101;
Matches 190; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MQSILMQAPLLIALGLLLIATPAQAHKKPSQSSFSWNCDEGKDPVIRSLTLEPDPPIV 60
 Db 1 MQSILMQAPLLIALGLLLIATPAQAHKKPSQSSFSWNCDEGKDPVIRSLTLEPDPPIV 60
 QY 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLIP 120
 Db 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLIP 120
 QY 121 TGSPCEPLRTYGLPCHCPKPGEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
 Db 121 TGSPCEPLRTYGLPCHCPKPGEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
 QY 181 LGCITKAASLKI 193
 Db 181 LGCITKAASLKI 193

RESULT 13

ADQ17712
 ID ADQ17712 standard; protein; 193 AA.
 AC ADQ17712;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Human soft tissue sarcoma-upregulated protein - SEQ ID 529.
 XX
 KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
 XX
 OS Homo sapiens.
 XX
 PN W02004048938-A2.
 XX
 PD 10-JUN-2004.
 XX
 PF 26-NOV-2003; 2003WO-US038193.
 XX
 PR 26-NOV-2002; 2002US-0429739P.
 XX
 PA (PROT-) PROTEIN DESIGN LABS INC.
 XX
 PI Aziz N, Ginsburg WM, Zlotnik A;
 XX
 DR WPI; 2004-441208/41.

Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.

Example 2; SEQ ID NO 529; 210pp; English.

The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated protein of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.

Sequence 193 AA;

Query Match 98.2%; Score 1000; DB 8; Length 193;
 Best Local Similarity 98.4%; Pred. No. 4.4e-101;
 Matches 190; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MQSILMQAPLLIALGLLLIATPAQAHKKPSQSSFSWNCDEGKDPVIRSLTLEPDPPIV 60
 Db 1 MQSILMQAPLLIALGLLLIATPAQAHKKPSQSSFSWNCDEGKDPVIRSLTLEPDPPIV 60
 QY 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLIP 120
 Db 61 PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLIP 120
 QY 121 TGSPCEPLRTYGLPCHCPKPGEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
 Db 121 TGSPCEPLRTYGLPCHCPKPGEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
 QY 181 LGCITKAASLKI 193
 Db 181 LGCITKAASLKI 193

RESULT 14

AAB31899
 ID AAB31899 standard; protein; 200 AA.
 AC AAB31899;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of a human protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine; ganglioside GM2 activator; saposin B; degenerative disease; glial cell; neurological disease; auto-immune disease; multiple sclerosis; toxicity; Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis; rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN W0200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIONERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Maicus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.

Detecting, preventing and treating degenerative, neurological and autoimmune diseases, particularly multiple sclerosis, using specified polypeptides or related nucleic acid or ligand.

Claim 1; Page 160-161; 209pp; French.

The present sequence represents a human protein, which is used in the method of the invention. The specification describes a method which uses at least one polypeptide or polynucleotide sequence belonging to the perlecan, precursor of the retinol-binding plasma protein, precursor of the ganglioside GM2 activator, calgranulin B or saposin B protein families. The method is used for detecting, preventing or treating a degenerative, neurological and/or auto-immune disease. The polynucleotides and polypeptides are used for diagnosis, prognosis, prevention and treatment of multiple sclerosis (in its various forms and phases). They may also be useful in cases of e.g. Alzheimer's and Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid polyarthritis and lupus erythematosus, including use as vaccines and in gene therapy (expression of sense or antisense sequences). They can also be used to assess efficacy of potential therapeutic agents, particularly compounds that reduce or inhibit toxicity towards glial cells

Sequence 200 AA;

Query Match 97.8%; Score 996; DB 4; Length 200;
Best Local Similarity 97.4%; Pred. No. 1.3e-100;
Matches 188; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MOSLMQAPLIIAGLLATPAQAHLKXPSQSSFSWDCPEGKDPVIRISLTLEPDPDIV 60
DB 8 MOSLMQAPLIIAGLLAQAQAHLKXPSQSSFSWDCDEGKDPVIRISLTLEPDPDIV 67

QY 61 PGNVTLISVWGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDVLDMLIP 120
DB 68 PGNVTLISVWGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDVLDMLIP 127

QY 121 TGBPCPEPLRTYGLPCHCPKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
DB 128 TGBPCPEPLRTYGLPCHCPKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKR 187

QY 181 LGCIKIAASLKI 193
DB 188 LGCIKIAASLKI 200

RESULT 15
AAB31900
ID AAB31900 standard; protein; 189 AA.
XX AC AAB31900;
XX AC AAB31900;
DT 15-MAY-2001 (first entry)
XX DE Amino acid sequence of a human protein.
XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX OS Homo sapiens.
XX WO200105422-A2.
XX 25-JAN-2001.
XX 17-JUL-2000; 2000WO-FR002057.
XX 15-JUL-1999; 99FR-00009372.
XX (INNR) BIOMERIEUX STELHYS.
XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX Claim 1; Page 161-162; 209pp; French.

CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly

CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 189 AA:
SQ

Query Match 97.0%; Score 987; DB 4; Length 189;
Best Local Similarity 98.9%; Pred. No. 1.1e-99;
Matches 187; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 MQAPLLIAGLLATPAQAHLKXPSQSSFSWDCPEGKDPVIRISLTLEPDPDIVVPGNV 64
DB 1 MQAPLLIAGLLATPAQAHLKXPSQSSFSWDCDEGKDPVIRISLTLEPDPDIVVPGNV 60

QY 65 TLSVWGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDVLDMLIPTGEP 124
DB 61 TLSVWGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDVLDMLIPTGEP 120

QY 125 CPEPLRTYGLPCHCPKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKRLGCI 184
DB 121 CPEPLRTYGLPCHCPKEGTYSLPKSEFVVPDLELPSWLTGNYRIESVLSGSKRLGCI 180

QY 185 KIAASLKI 193
DB 181 KIAASLKI 189

Search completed: November 17, 2004, 06:12:06
Job time : 32.4237 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 05:44:53 ; Search time 7.65624 Seconds
(without alignments)
1671.758 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 1018

Sequence: 1 MQSLMQAPLLIALGLLLATP.....LSSSGKRLGCIKIAASLKGI 193

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pap:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pap:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pap:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pap:*
5: /cgn2_6/ptodata/1/iaa/6C.COMB.pap:*
6: /cgn2_6/ptodata/1/iaa/6D.COMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1005	98.7	193	4	US-09-183-841-1
2	857	84.2	178	4	US-09-183-841-2
3	88.5	8.7	310	4	US-09-976-594-807
4	86.5	8.5	410	3	US-08-630-172-17
5	86.5	8.5	410	3	US-09-375-419-17
6	86.5	8.5	768	3	US-09-302-812-8
7	86.5	8.5	768	3	US-09-511-477-8
8	86.5	8.5	768	3	US-09-511-507-8
9	86	8.4	143	4	US-09-860-793-3
10	83	8.2	572	6	5256770-7
11	82	8.1	446	1	US-08-307-444A-5
12	82	8.1	446	1	US-08-587-389-5
13	82	8.1	456	1	US-08-307-444A-3
14	82	8.1	456	1	US-08-307-444A-4
15	82	8.1	456	1	US-08-587-389-3
16	82	8.1	456	1	US-08-587-389-4
17	82	8.1	475	1	US-08-307-444A-1
18	82	8.1	475	1	US-08-307-444A-2
19	82	8.1	475	1	US-08-587-389-1
20	82	8.1	475	1	US-08-587-389-2
21	82	8.1	476	1	US-08-014-723-1
22	82	8.1	476	1	US-08-014-723-2
23	82	8.1	476	1	US-08-014-723-18
24	82	8.1	476	1	US-08-110-011A-1
25	82	8.1	476	1	US-08-110-011A-2
26	82	8.1	476	1	US-08-110-011A-18
27	82	8.1	494	1	US-08-014-723-14

28	82	8.1	494	1	US-08-014-723-16
29	82	8.1	494	1	US-08-110-011A-14
30	82	8.1	494	1	US-08-110-011A-16
31	82	8.1	497	1	US-08-312-870-3
32	82	8.1	497	4	US-09-331-793-4
33	82	8.1	498	2	US-08-733-564-2
34	82	8.1	575	1	US-08-261-206A-59
35	82	8.1	575	1	US-08-312-870-1
36	82	8.1	575	1	US-08-170-290A-54
37	82	8.1	575	4	US-09-880-484D-2
38	82	8.1	575	4	US-10-438-648-2
39	82	8.1	575	6	5466668-6
40	82	8.1	746	4	US-09-370-838-185
41	82	8.1	746	4	US-09-854-133-185
42	81	8.0	302	4	US-09-248-796A-14539
43	78.5	7.7	611	2	US-08-677-049-2
44	77.5	7.6	126	4	US-09-860-793-1
45	77.5	7.6	732	1	US-08-317-522A-5

ALIGNMENTS

RESULT 1
US-09-183-841-1
; Sequence 1, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183.841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (33)..(55)
; FEATURE:
; OTHER INFORMATION: residues 56-63 are included in a further precursor
; OTHER INFORMATION: form of the protein
US-09-183-841-1

Query Match		98.7%	Score 1005;	DB 4;	Length 193;
Best Local Similarity		99.0%	Pred. No. 5.1e-110;		
Matches 191;		Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	1	MQSLMQAPLLIALGLLLATPAQAHLKPKS	QSLSSFSWNCNCEGKDPVIRSLTLEPDPV	60	
Db	1	MQSLMQAPLLIALGLLLATPAQAHLKPKS	QSLSSFSWNCNCEGKDPVIRSLTLEPDPV	60	
Qy	61	PGNVTLSVWGTSVPLSLKVDLVLEKEVAGL	MIKIPCTDYIGSCFTEHFCVDLMLP	120	
Db	61	PGNVTLSVWGTSVPLSLKVDLVLEKEVAGL	MIKIPCTDYIGSCFTEHFCVDLMLP	120	
Qy	121	TGECPCPEPLRTYGHPCPCPKEGTYSLPK	SEFAVDDLELPSWLTGNYRIBSVLSSGKR	180	
Db	121	TGECPCPEPLRTYGHPCPCPKEGTYSLPK	SEFAVDDLELPSWLTGNYRIBSVLSSGKR	180	
Qy	181	LGCIKIAASLKGI	193		
Db	181	LGCIKIAASLKGI	193		

RESULT 2
US-09-183-841-2
; Sequence 2, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:

```

; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680e1 Inhibitor of Platelet Activating Factor
; FILE REFERENCE: van0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: His tag at residues 1 to 17
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: amino acid
; OTHER INFORMATION: sequence of GM2 protein using His6 tag
US-09-183-841-2

Query Match      84.2%; Score 857; DB 4; Length 178;
Best Local Similarity 98.8%; Pred. No. 1.2e-92;
Matches 160; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 32 SFSWNCPEGKDPKPAVIRSLTLEPDPPIVPGNVTLSVWGSTVPLSPKVDLVLEKEVA 91
Db 17 SFSWNCDEGKDPKPAVIRSLTLEPDPPIVPGNVTLSVWGSTVPLSPKVDLVLEKEVA 76
QY 92 GLWIKIPCTDYIGSCTFEHFCDDVLDMLIPTGEPCEPELRTYGLPCHCPFKEGTYSLPKSE 151
Db 77 GLWIKIPCTDYIGSCTFEHFCDDVLDMLIPTGEPCEPELRTYGLPCHCPFKEGTYSLPKSE 136
QY 152 FAVPDLPLSWLTGNYRIEVSLSGKRLGCIKIAASLKI 193
Db 137 FVVPDLPLSWLTGNYRIEVSLSGKRLGCIKIAASLKI 178

RESULT 3
US-09-376-594-807
; Sequence 807, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 807
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 2770104CD1
US-09-376-594-807

Query Match      8.7%; Score 88.5; DB 4; Length 310;
Best Local Similarity 24.7%; Pred. No. 0.075;
Matches 55; Conservative 20; Mismatches 67; Indels 81; Gaps 10;

QY 9 LLIALGLLATPAQAH---LKKPSQLSFSWNCDEGKDPKPAVIRSLTLEPDPPIVPGN-V 64
Db 51 VLICGLSLGPRTHVQGTIPKP-----TLWAEPSVITQSPV 89
QY 65 TLSVWGSTVPLSPKVDLVLEKEVAGLWKIPCTD-----YIGSCTFEH----- 110
Db 90 TLSQCSGLEAQ-----EYRLYREKKAS-WITRIRPELVNKGQPHISITWHTGRVGCQ 143
QY 111 -----PCDVLDMLI-----PTGEPCEPELRTYG---LPCHCPFKEGTYSLPKSE 151

```

```

Db 144 YYSRARWSSELDPLVAGDDRSYQNPISQSPSPGVVTPGKNVTLLCQSGQFHFLTKEG 203
QY 152 FAVPDLLELPS-----WLTGNYRIEVSLS 177
Db 204 AGHPPLHLRSEHQAOQQAQFRMGPTSAHVGTIRCYSSLSN 246

RESULT 4
US-08-630-172-17
; Sequence 17, Application US/08630172
; Patent No. 6060054
; GENERAL INFORMATION:
; APPLICANT: Staerz, Uwe
; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
; TITLE OF INVENTION: LYMPHOCYTE VETO
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, 35th Floor
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,172
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2879-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-630-172-17

Query Match      8.5%; Score 86.5; DB 3; Length 410;
Best Local Similarity 25.6%; Pred. No. 0.19;
Matches 42; Conservative 14; Mismatches 43; Indels 65; Gaps 9;

QY 26 KKPSSQLSFS-----WDCNCFEGKDPKPAVIRSLTLEPDPPIV-----PGNVTLSVWGST- 72
Db 46 KGFSLNDRADRSRLWD---QGNFPLIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTA 102
QY 73 -----SVPLSSP-----LKVDLVLEKEVAGLWKIP 98
Db 103 NSDTHLLQGOSLFLTESPPGSPSVOCRSRPNKIQGGTSLVS-QLELDAGTWI--- 158
QY 99 CTDYIGSCTFEHFCDDVLDMLIPTG---EPCPEPLRTYGLPCHCP 139
Db 159 CTVLQKQKVEFKIDIVLAEPRGPTIKPCP-----PKCP 194

RESULT 5
US-09-375-419-17
; Sequence 17, Application US/09375419
; Patent No. 6264950
; GENERAL INFORMATION:
; APPLICANT: Staerz, Uwe
; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
; TITLE OF INVENTION: LYMPHOCYTE VETO
; NUMBER OF SEQUENCES: 41

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/375,419
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,172
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-375-419--17

Query Match      8.5%   Score 86.5; DB 3; Length 410;
Match Local Similarity 25.6%; Pred. No. 0.19; 43; Indels 65; Gaps 9;
Matches 42; Conservative 14; Mismatches 14;

QY 26 KPSQLSSFS-----WDNCFEGKDPVAVIRSLTLEPPIVV-----PGNVTLSVVGST- 72
Db 46 KGFSLNDRADRRSLWD---QGNFPLIKNKLESDTYICEVEDQKEVQLLVFLTA 102
QY 73 -----SVPLSSP-----LKVDLVLEKEVAGLWIKIP 98
Db 103 NSDTHLLQGQSLATLTLESPPGSSPSVQCRSPRGKNIQGGKTLVS-QLELQDAGTWI--- 158
QY 99 CTDYIGSCTEFHCDVLDMLPTG---EPCPEPLRTYGLPCHCP 139
Db 159 CTVLQNKQKVEFKIDIVVLAEPRGPTIKPCP-----PCKCP 194

RESULT 6
US-09-302-812-8
; Sequence 8, Application US/09302812B
; Patent No. 633148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPHYDROLASE (PARG) ENZY
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,444A
; FILING DATE: 19-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/835,436
; FILING DATE: 26-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: OLIFF, JAMES A.
; REGISTRATION NUMBER: 27,075
; REFERENCE/DOCKET NUMBER: JAO 27706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6400
; TELEFAX: (703) 836-2787
; TELEX: 90-1799 PTO ALEX
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-307-444A-5

Query Match      8.1%; Score 82; DB 1; Length 446;
Best Local Similarity 23.3%; Pred. No. 0.74;
Matches 49; Conservative 23; Mismatches 74; Indels 64; Gaps 12;

QY 8 PLLALGLLLAT-----PAQHLKKPSLSSFSWDCFEKDPVIRSLTLEPDPV 59
Db 117 PLCVAVSAAEATVPSEPIWEQCEVADGFLCEHF-----PATCRPLAVEPGAAA 168
QY 60 VPGNVTLVS-----VGSTSVPLSSPLKVDLVL-----EKEVAGLWI 95
Db 169 AAVSITYGTPFAARGADFOALPVSSAA--VAPLGLQLMCTAPPGAVGHWAREAPGAW- 225
QY 96 KIPCTDYIGSTFHFCDVLDMLIPTGEPCEP-----LRTYGLPCHCFKGTYSLPKSE 151
Db 226 --DCSVENGCG--EHACNA----IPGAPRCQPAGAAALQADGRSCTASATQSCNDLCE-H 276
QY 152 FAVPDLEPLSWLTGNYRIEVSLSGKRL 181
Db 277 FCVENPDQP-----GSY---SCMCETGYRL 298

RESULT 12
US-08-587-389-5
; Sequence 5, Application US/08587389
; Patent No. 5695964
; GENERAL INFORMATION:
; APPLICANT: NII, ATSUSHI
; APPLICANT: MORISHITA, HIDEAKI
; APPLICANT: UEMURA, AKIO
; APPLICANT: MOCHIDA, EI
; TITLE OF INVENTION: TRUNCATED THROMBOMODULIN, RECOMBINANT
; TITLE OF INVENTION: PRODUCTION THEREOF, AND THERAPEUTIC AGENT (AS AMENDED)
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/587,389
; FILING DATE: 19-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

;
; FILING DATE: 17-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/307,444
; FILING DATE: 19-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: OLIFF, JAMES A.
; REGISTRATION NUMBER: 27,075
; REFERENCE/DOCKET NUMBER: JAO 27706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6400
; TELEFAX: (703) 836-2787
; TELEX: 90-1799 PTO ALEX
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-587-389-5

Query Match      8.1%; Score 82; DB 1; Length 446;
Best Local Similarity 23.3%; Pred. No. 0.74;
Matches 49; Conservative 23; Mismatches 74; Indels 64; Gaps 12;

QY 8 PLLALGLLLAT-----PAQHLKKPSLSSFSWDCFEKDPVIRSLTLEPDPV 59
Db 117 PLCVAVSAAEATVPSEPIWEQCEVADGFLCEHF-----PATCRPLAVEPGAAA 168
QY 60 VPGNVTLVS-----VGSTSVPLSSPLKVDLVL-----EKEVAGLWI 95
Db 169 AAVSITYGTPFAARGADFOALPVSSAA--VAPLGLQLMCTAPPGAVGHWAREAPGAW- 225
QY 96 KIPCTDYIGSTFHFCDVLDMLIPTGEPCEP-----LRTYGLPCHCFKGTYSLPKSE 151
Db 226 --DCSVENGCG--EHACNA----IPGAPRCQPAGAAALQADGRSCTASATQSCNDLCE-H 276
QY 152 FAVPDLEPLSWLTGNYRIEVSLSGKRL 181
Db 277 FCVENPDQP-----GSY---SCMCETGYRL 298

RESULT 13
US-08-307-444A-3
; Sequence 3, Application US/08307444A
; Patent No. 5516659
; GENERAL INFORMATION:
; APPLICANT: NII, ATSUSHI
; APPLICANT: MORISHITA, HIDEAKI
; APPLICANT: UEMURA, AKIO
; APPLICANT: MOCHIDA, EI
; TITLE OF INVENTION: ANTICOAGULANT POLYPEPTIDES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,444A
; FILING DATE: 19-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: US 07/835,436
FILING DATE: 26-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: OLIFF, JAMES A.
REGISTRATION NUMBER: 27,075
REFERENCE/DOCKET NUMBER: JAO 27706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6400
TELEFAX: (703) 836-2787
TELEX: 90-1799 PTO ALEX
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-307-444A-3

Query Match 8.1%; Score 82; DB 1; Length 456;
Best Local Similarity 23.3%; Pred. No. 0.77;
Matches 49; Conservative 23; Mismatches 74; Indels 64; Gaps 12;
QY 8 PLLIALGLLIAT-----PAQAHLLKPSQLSFSWDCNCFEGKDPVAVIRSLTLEPPIV 59
Db 117 PLCVAVSAEATVPSEPIWEEQQCEVKADGFLCFEHP-----PATCRPLAVEPGAAA 168
QY 60 VPGNVTLV-----VGSTSVPLSSPLKVDLVL-----EKEVAGLWI 95
Db 169 AAVSIITYGTFFAARGADFOALPVGSSAA--VAPLGLQLMCTAPPGAVQGHWAREAPGAW- 225
QY 96 KIPCTDYIGCTEHCDFVDMLIPTGCPCEP-----LRTYGLPCHCPFKEGTYSLPKSE 151
Db 226 --DCSVENGCC--EHACNA----IPGAPRCQCPAGALQADGRSCTASATQSCNDLCE-H 276
QY 152 FAYPDLEPLSWLTGNYRIEVSLSGKRL 181
Db 277 FCVNPDPQ-----GSY---SCMCETGYRL 298

RESULT 14
US-08-307-444A-4
Sequence 4, Application US/08307444A
Patent No. 5516659
GENERAL INFORMATION:
APPLICANT: NII, ATSUSHI
APPLICANT: MORISHITA, HIDEAKI
APPLICANT: UEMURA, AKIO
APPLICANT: MOCHIDA, EI
TITLE OF INVENTION: ANTICOAGULANT POLYPEPTIDES
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/307,444A
FILING DATE: 19-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/835,436
FILING DATE: 26-FEB-1992
ATTORNEY/AGENT INFORMATION:

NAME: OLIFF, JAMES A.
REGISTRATION NUMBER: 27,075
REFERENCE/DOCKET NUMBER: JAO 27706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6400
TELEFAX: (703) 836-2787
TELEX: 90-1799 PTO ALEX
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-307-444A-4

Query Match 8.1%; Score 82; DB 1; Length 456;
Best Local Similarity 23.3%; Pred. No. 0.77;
Matches 49; Conservative 23; Mismatches 74; Indels 64; Gaps 12;
QY 8 PLLIALGLLIAT-----PAQAHLLKPSQLSFSWDCNCFEGKDPVAVIRSLTLEPPIV 59
Db 117 PLCVAVSAEATVPSEPIWEEQQCEVKADGFLCFEHP-----PATCRPLAVEPGAAA 168
QY 60 VPGNVTLV-----VGSTSVPLSSPLKVDLVL-----EKEVAGLWI 95
Db 169 AAVSIITYGTFFAARGADFOALPVGSSAA--VAPLGLQLMCTAPPGAVQGHWAREAPGAW- 225
QY 96 KIPCTDYIGCTEHCDFVDMLIPTGCPCEP-----LRTYGLPCHCPFKEGTYSLPKSE 151
Db 226 --DCSVENGCC--EHACNA----IPGAPRCQCPAGALQADGRSCTASATQSCNDLCE-H 276
QY 152 FAYPDLEPLSWLTGNYRIEVSLSGKRL 181
Db 277 FCVNPDPQ-----GSY---SCMCETGYRL 298

RESULT 15
US-08-587-389-3
Sequence 3, Application US/08587389
Patent No. 5695964
GENERAL INFORMATION:
APPLICANT: NII, ATSUSHI
APPLICANT: MORISHITA, HIDEAKI
APPLICANT: UEMURA, AKIO
APPLICANT: MOCHIDA, EI
TITLE OF INVENTION: TRUNCATED THROMBOMODULIN, RECOMBINANT
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/587,389
FILING DATE: 17-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/307,444
FILING DATE: 19-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: OLIFF, JAMES A.
REGISTRATION NUMBER: 27,075

Fri Nov 19 14:12:54 2004

```

; REFERENCE/DOCKET NUMBER: JAO 27706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6400
; TELEFAX: (703) 836-2787
; TELEX: 90-1799 PTO ALEX
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-587-389-3

Query Match      8.1%; Score 82; DB 1; Length 456;
Best Local Similarity 23.3%; Pred. No. 0.77; 74; Indels 64; Gaps 12;
Matches 49; Conservative 23; Mismatches 74; Indels 64; Gaps 12;

QY      8 PLLIALGLLLAT-----PAQHLKPKSQLSFSWDNCFEGKDPVIRSLTLEPDPIV 59
Db      117 PLCVAVSAEATVPSEPIWEEQCEVKADGFLCFEHE-----PATCRPLAVEPGAAA 168

QY      60 VPGNVTLSV-----YGSTSVPLSSPLKVDLVL-----EKEVAGLWI 95
Db      169 AAVSITVGTFFAARGADFOALPVGSSAA--VAPLGLQLMCTAPPVAVQGHWAREAPGAW- 225

QY      96 KIPCTDYIGSCTEHFCDVLDMLIPTGCPCEP-----LRTYGLPCHCFKFKEGTYSLPKSE 151
Db      226 --DCSVENGGC--EHACNA----IPGAPRCQCPAGALQADGRSCTASATQSCNDLCE-H 276

QY      152 FAVPDLELPSSLTGTGNVRIESVLSGKRL 181
Db      277 FCVENPDQP-----GSY---SCMCETGYRL 298

```

Search completed: November 17, 2004, 06:43:24
Job time : 8.70624 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 05:48:19 ; Search time 25.0941 Seconds
(without alignments)
2723.613 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 1018

Sequence: 1 MQSLMQAPLLIALGLLTP.....LSSGKRLGCIKIAASLKG1 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1005	98.7	193	14	US-10-170-385-389
2	412	40.5	191	15	US-10-264-049-2611
3	353	34.7	61	9	US-09-864-761-34809
4	162.5	16.0	126	10	US-09-764-891-4977
5	98.5	9.7	273	17	US-10-425-115-297138
6	91.5	9.0	641	14	US-10-369-493-20746
7	90	8.8	651	14	US-10-369-493-20109
8	88	8.6	796	16	US-10-437-963-187458
9	86.5	8.5	196	17	US-10-425-115-251307
10	86.5	8.5	574	17	US-10-725-013-2
11	86.5	8.4	768	9	US-09-973-451-8
12	86	8.4	143	9	US-09-860-793-3
13	86	8.4	474	14	US-10-369-493-19074

14	86	8.4	575	15	US-10-094-886-196	Sequence 196, App
15	86	8.4	611	14	US-10-369-493-7417	Sequence 7417, App
16	86	8.4	633	14	US-10-369-493-4658	Sequence 4658, App
17	86	8.4	2293	16	US-10-032-438B-2	Sequence 2, Appli
18	84	8.3	490	16	US-10-437-963-180789	Sequence 180789,
19	83.5	8.2	132	15	US-10-424-599-255741	Sequence 255741,
20	83	8.2	688	15	US-10-282-122A-47772	Sequence 47772, A
21	82.5	8.1	342	17	US-10-425-115-220322	Sequence 220322,
22	82.5	8.1	342	17	US-10-425-115-220323	Sequence 220323,
23	82.5	8.1	352	15	US-10-425-114-63175	Sequence 63175, A
24	82.5	8.1	371	15	US-10-425-114-62424	Sequence 62424, A
25	82.5	8.1	372	15	US-10-425-114-53454	Sequence 53454, A
26	82.5	8.1	372	15	US-10-425-114-72059	Sequence 72059, A
27	82.5	8.1	375	15	US-10-425-114-58493	Sequence 58493, A
28	82.5	8.1	792	14	US-10-309-422-38	Sequence 38, Appli
29	82.5	8.1	824	16	US-10-467-595-4	Sequence 4, Appli
30	82.5	8.1	957	14	US-10-309-422-36	Sequence 36, Appli
31	82.5	8.1	957	14	US-10-309-422-14	Sequence 14, Appli
32	82.5	8.1	958	14	US-10-309-422-26	Sequence 26, Appli
33	82.5	8.1	1091	14	US-10-309-422-12	Sequence 12, Appli
34	82.5	8.1	1092	14	US-10-309-422-24	Sequence 24, Appli
35	82	8.1	497	14	US-10-298-796-4	Sequence 4, Appli
36	82	8.1	575	9	US-09-938-405-2	Sequence 2, Appli
37	82	8.1	575	10	US-09-880-464-2	Sequence 2, Appli
38	82	8.1	575	14	US-10-150-440-3	Sequence 3, Appli
39	82	8.1	575	14	US-10-438-648-2	Sequence 2, Appli
40	82	8.1	575	15	US-10-410-195-2	Sequence 2, Appli
41	82	8.1	575	15	US-10-373-801-29	Sequence 29, Appli
42	82	8.1	631	14	US-10-150-440-1	Sequence 1, Appli
43	82	8.1	631	16	US-10-741-601-309	Sequence 309, App
44	82	8.1	651	14	US-10-369-493-4736	Sequence 4736, App
45	82	8.1	659	14	US-10-369-493-7495	Sequence 7495, App

ALIGNMENTS

RESULT 1
US-10-170-385-389
; Sequence 389, Application US/10170395
; Publication No. US20030203372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Jonathan
; APPLICANT: White, Katie Mary
; APPLICANT: Binley, William Nigel
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 53268200100
; CURRENT APPLICATION NUMBER: US/10/170.385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 389
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-170-385-389

Query Match 98.7%; Score 1005; DB 14; Length 193;
Best Local Similarity 99.0%; Pred. No. 1.5e-96;
Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQSLMQAPLLIALGLLTPAQAHKPKSLSFSDNCFCGKDPVIRSLTLEPDIW 60

```

Db      1  MQSLMQAPLIIAGLLIATPAQHLKPKSLSFSDNCDGKDPVAVIRSLTLEPDPV 60
QY      61  PGNVTLVSGVSTVPLSSPLKVDLLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLIP 120
Db      61  PGNVTLVSGVSTVPLSSPLKVDLLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLIP 120
QY      121  TGPCCPEPLRTYGLPCHCPKEGTYSLPKSEFAVDPDLPSWLTGNGYRTIESVLSGGR 180
Db      121  TGPCCPEPLRTYGLPCHCPKEGTYSLPKSEFAVDPDLPSWLTGNGYRTIESVLSGGR 180
QY      181  LGCIKIAASLKGI 193
Db      181  LGCIKIAASLKGI 193

```

RESULT 2

```

US-10-264-049-2611
; Sequence 2611, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/19569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 2611
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (141)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (142)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (184)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (187)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2611

```

```

Query Match      40.5%; Score 412; DB 15; Length 191;
Best Local Similarity 48.4%; Pred. No. 1.3e-34;
Matches 78; Conservative 24; Mismatches 51; Indels 8; Gaps 2;
QY      13  LGLLIA-----TPAQLKPKSLSFSDNCDGKDPVAVIRSLTLEPDPVPGVNT 65
Db      24  LGLLIAGPAHAHVPAHVPVNPQVIFSFENCHERKDPVLKSMTLEPDPPIAIPGNT 83
QY      66  LSVGSTVPLSSPLKVDLLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMLITGAPC 125
Db      84  ISAEIQVRVPLSSPKVELIIEKKVANTWIKVPCMSHV-RCIFEDICQLDPLPPGQXX 142
QY      126  PEPLTYGLPCHCPKEGTYSLPKSEFAVDPDLPSWLTG 166
Db      143  PEPLTYGLPCTVPSSRLNLAQLKLPCTNDLPGLITSG 183

```

RESULT 3

```

US-09-864-761-34809

```

```

; Sequence 34809, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34809
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011342.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99
; OTHER INFORMATION: EST HUMAN HIT: BE182886.1, EVALUATE 9.00e-34
; OTHER INFORMATION: SWISSPROT HIT: P17900, EVALUATE 1.00e-34
US-09-864-761-34809

```

```

Query Match      34.7%; Score 353; DB 9; Length 61;
Best Local Similarity 100.0%; Pred. No. 4.4e-29;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 82 VDLVLEKEVAGLWKIPCTDYIGSTGEHFCVDVLDMLIPTGECPEPLRTYGLPCHCPFX 141
 Db 1 VDLVLEKEVAGLWKIPCTDYIGSTGEHFCVDVLDMLIPTGECPEPLRTYGLPCHCPFX 60

Qy 142 E 142
 Db 61 E 61

RESULT 4
 US-09-764-891-4977
 ; Sequence 4977, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4977
 ; LENGTH: 126
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (119)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (122)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (123)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-764-891-4977

Query Match 16.0%; Score 162.5; DB 10; Length 126;
 Best Local Similarity 54.2%; Pred. No. 9.1e-09;
 Matches 32; Conservative 7; Mismatches 13; Indels 7; Gaps 1;

Qy 13 LGLLLA-----TPAQLKKPQLSFSWDCNCFEGKDPVIRSLTLEPPPIVPGNV 64
 Db 24 LGLLLAGAAHAHVPAHPNPFPQVTSFFWENCHERKDPVLLKSMTEPDPIAYPGNV 82

RESULT 5
 US-10-425-115-297138
 ; Sequence 297138, Application US/10425115
 ; Publication No. US20040214272A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53222)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369326
 ; SEQ ID NO 297138
 ; LENGTH: 273
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: MFT4577_34069C.1.pap
 US-10-425-115-297138
 Query Match 9.7%; Score 98.5; DB 17; Length 273;
 Best Local Similarity 26.7%; Pred. No. 0.12;

Matches 59; Conservative 24; Mismatches 73; Indels 65; Gaps 14;
 Qy 3 SLMQAPL---LIALGLLLATPAQHLKK-----PSQLSFSWDCNCFEGKDPVIRSLTLEPPPIVPGNV 50
 Db 70 ALLPAALPASILFPAASILYSPLOQAVLPRCGISLCPAELAPA--PSCCSSRRPAPLAWSS 127
 Qy 51 LTLEPDPPIVPGNVTLTVSGSTSVPLSSPLKV-----DLVLEKEVAGLW----- 94
 Db 128 SFLVFEVPCVFP--VTSMAPFSL--PLSSPLRVPLVSLRRARSFLCVPAAATSLWCFLLPA 184
 Qy 95 -IKIPCTDYIGSTGEHFCVDVLDMLIPTGECPEPLRTYGLPCH-----CPFKECTY 145
 Db 185 RAKFPCSLALGPC-----STVPCFSARV-KEPCRVCLGRKPVCPRR--AC 227
 Qy 146 SLPKSEFAVPDLPLSLWLTGNYRIESVLSSSG--KRLGCI 184
 Db 228 CSPKR---PMLQRPYFMSFHVGVSVSSSGINKQTGAV 264

RESULT 6
 US-10-369-493-20746
 ; Sequence 20746, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 20746
 ; LENGTH: 641
 ; TYPE: PRT
 ; ORGANISM: Rhodospseudomonas palustris
 US-10-369-493-20746

Query Match 9.0%; Score 91.5; DB 14; Length 641;
 Best Local Similarity 25.4%; Pred. No. 2.1;
 Matches 47; Conservative 30; Mismatches 63; Indels 45; Gaps 11;

Qy 45 PAVIRSLTLEPD--PIVVPGNVTLTVSGSTSVPLS-----SPLKYDLVLEK 88
 Db 351 PTAIRALMQAGDEPVKTSRKLGLSGVGEPIINPEAWVHRVVGEDRCFI-VDTWWQT 409
 Qy 89 EVAGLMI-KIPCTDYI--GSCTFEHFCVDVLDMLIP-----TGEPC-----PEPLRT 131
 Db 410 ETGGILITPLPGATKLPKGSATRFFGVVPEILDPEGNVLEGECTGNLCARSFPGQMT 469
 Qy 132 -YGLPCHCPKPGYSLPKSEFAV-----PDLEPLSLWLTGNYRIESVLSSSGKRLGCIK 185
 Db 470 VYG--DHARFEQYFSAKGYKGYFTGDCGRDRTDGYMITG---RVDDVINVSGHRMGTA 524
 Qy 186 IAASL 190
 Db 525 VESSL 529

RESULT 7
 US-10-369-493-20109
 ; Sequence 20109, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.

Query Match	8.6%;	Score 88;	DB 16;	Length 796;
Best Local Similarity	24.2%;	Pred. No. 6.3;		
Matches	57;	Conservative 22;	Mismatches 63;	Indels 94; Gaps 12;
Qy	24	HLKPSQLSFSWNCFEFGKDPVIRSLTLEPDPV----	VPGNVTL-----	66
Db	108	NLKKAAS-----DNEFTGKLPDLVLSLTELEDVLNCRISENLVTFDSKFAALTWLF	160	
Qy	67	-----SVVSGTSVPGLSKV--DLVLEKEVAG-----		92
Db	161	LGNNSLIGTLTPDVISSSLKVMNLVANNILVLSGTSKNSDISTRGSNTIYEADATNLGDASY	220	
Qy	93	-----LWIKIPCTDYIGSCT-----FEHFCVDILD-MLIPTGEPCCPEPLRTYGLFCH	137	

```

Query Match      8.5%; Score 86.5; DB 17; Length 574;
Best Local Similarity 24.1%; Pred. No. 5.9;
Matches 51; Conservative 21; Mismatches 71; Indels 69; Gaps 13;

QY 8 PLLIALGLLLAT-----PAQAHKKPSQSSFNDCNCFEGKDPVAVIRSLTLEPDPV 59
Db 135 PLCVAVSAAEATVSEPIWEQQCEVKADGLCFSEHF-----PATCRPLAVEPGAA 186
QY 60 VPGNVTLSV-----VGSTVPLSSPLKVDLVL-----EKEVAGLWI 95
Db 187 AAVSITVGTPEAARGADFOALPVGSSAA--VAPLGLQLMCTAPPAGVQCHWAREAPGAW- 243
QY 96 KIPCTDYIGCTFHFCDVLDMLIPTGEPCEP-----LRTYGLPCHCFKSGTSLPK-- 149
Db 244 --DCSVENGCC--EHACNA-----IPGARPCQCPAGALQADGRSC-----TASTQSCNDLC 291
QY 150 SEFAPVPLEPDLPSWLTGTGNRYIESVLSSSGKRL 181
Db 292 EHFVCPNPDP-----GSY---SCMCEITGYRL 315

RESULT 11
US-09-973-451-8
; Sequence 8, Application US/09573451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES.
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-973-451-8

Query Match      8.5%; Score 86.5; DB 9; Length 768;
Best Local Similarity 25.4%; Pred. No. 8.7;
Matches 53; Conservative 25; Mismatches 94; Indels 37; Gaps 9;

QY 3 SLMOAPLLIALGLLLATPAQAHKKPSQSSFNDCNCFEGKDPVAVIRSLTLEPDPV 51
Db 245 SQQQISCLANAFCTFPRRWLTKKSYSTFPDINFNRLLYQSTGPVLEKLCIMHYFR 304
QY 52 ---TLEPDPVIVPGNVTLSVVGSS-----TSVPLSS--PLKVDL--VLEKEVAGLW 94
Db 305 RVCPTERDASNVPTGVTVFVRSGLPEHLIDWSQSAALPGDVPLHVDAGETIEDEGILL 364
QY 95 IKIPCTDYIGCTFHFCDVLDMLIPTGEPCEPPL--RTYGLPCHCFKSGTSLPKSE 151
Db 365 QVDFANKYLGGVGLGHGCVQEEIRFVI-----CPDLLVGKLFTECLRPF--EALVLMGAER 418
QY 152 FAVPDLPLPSWLTGTGNRYIESVLSSSGKR 180
Db 419 YSNYTGAGSFEWSGNFEDSTPRDSSGR 447

RESULT 12
US-09-860-793-3
; Sequence 3, Application US/09860793

```

```

; Patent No. US20020136734A1
; GENERAL INFORMATION:
; APPLICANT: Pruett, John H
; APPLICANT: Temeyer, Kevin B
; APPLICANT: Kunz, Sidney E
; APPLICANT: Fisher, William F
; TITLE OF INVENTION: Vaccines for the Protection of Cattle from Psoroptic
; TITLE OF INVENTION: Scabies
; FILE REFERENCE: Docket 0047.96 - John H. Pruett et al.
; CURRENT APPLICATION NUMBER: US/09/860,793
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/366,603
; PRIOR FILING DATE: 1999-08-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Psoroptes ovis
US-09-860-793-3

Query Match      8.4%; Score 86; DB 9; Length 143;
Best Local Similarity 22.8%; Pred. No. 1.1;
Matches 43; Conservative 32; Mismatches 56; Indels 58; Gaps 10;

QY 10 LIALGLLLATPAQAHKKPSQSSFNDCNCFEGKDPVAVIRSLTLE---PDPVIVPGNVTL 66
Db 5 LWLAITLAVVSAGVK-----FQDCGKE-----VESLEVGCSGDYCVIHKGKKL 51
QY 67 SV-VGSTVPLSSPLKVDLLEKEVAGLWIKIPCTDYIGCTFHFCDVLDMLIPTGEP 125
Db 52 DLAISVTNSQDSANLKLDIV--ADINGVQIIEVPGVDHGD----- 88
QY 126 PEPLRTYGLPCH---CPFKETSYLPKSEFAVPDLEPFLSWLTGTGNRYIESVLSSSGKRIG 182
Db 89 -----CHYVKCPKIKGQHFVDVKYTSIPAL-LP---TTRAKIILAKIIGDKELG-G 133
QY 183 CIKTAASLK 191
Db 134 CIVINGRIQ 142

RESULT 13
US-10-369-493-19074
; Sequence 19074, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19074
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Anabaena PCC7120
US-10-369-493-19074

Query Match      8.4%; Score 86; DB 14; Length 474;
Best Local Similarity 27.2%; Pred. No. 5.2;
Matches 41; Conservative 22; Mismatches 50; Indels 38; Gaps 10;

QY 68 VVGSTVPLSSPLKVDLLEKEVAGLWIKIPCTDYI---GSCTFEH---PCDVLDM--- 117
Db 228 VIGGERCPI-----VDTWWQETGTGIMI-TPLPGAI STKPGSATLPPFGILADIVDLEGN 281

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 05:43:48 ; Search time 5.41009 Seconds
(without alignments)
3432.447 Million cell updates/sec

Title: US-10-030-937-9
Perfect score: 1018
Sequence: 1 MQSLMQAPLIALGLLLATP.....LSSSGKRLGCIKTAASLKGI 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: Pirl:*
2: Pirl2:*
3: Pirl3:*
4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1005	98.7	193	2 154178	ganglioside M2 act
2	996	97.8	200	2 S22411	ganglioside M2 act
3	844	82.9	162	2 S13195	ganglioside M2 act
4	738.5	72.5	193	2 S35613	ganglioside M2 act
5	113	11.1	152	2 T49126	hypothetical prote
6	94	9.2	171	2 G72563	hypothetical prote
7	92	9.0	294	2 T34048	hypothetical prote
8	86.5	8.5	1060	2 A10201	beta-galactosidase
9	82	8.1	575	1 THHUB	zinc-finger protei
10	81	8.0	2180	2 A47651	thrombomodulin pre
11	80.5	7.9	504	2 AD3629	vacc protein [limpo
12	79	7.8	353	2 T49440	alcohol dehydrogen
13	79	7.8	456	2 T71402	probable glucosylt
14	78.5	7.7	583	2 F69153	conserved hypothet
15	78.5	7.7	693	2 T25878	hypothetical prote
16	78.5	7.7	708	2 T22377	hypothetical prote
17	78	7.7	551	2 AC2311	hypothetical prote
18	77.5	7.6	778	2 T38487	tastin - human
19	77	7.6	20	2 S56005	lysosomal protein
20	77	7.6	283	2 D69743	hypothetical prote
21	77	7.6	463	2 T21042	hypothetical prote
22	76.5	7.5	524	2 AF3096	conserved hypothet
23	76.5	7.5	524	2 C98150	hypothetical prote
24	76.5	7.5	819	2 B87386	hypothetical prote
25	76	7.5	485	2 T33865	hypothetical prote
26	75.5	7.4	249	2 A99222	conserved hypothet
27	75.5	7.4	894	2 T15769	hypothetical prote
28	75	7.4	320	2 JCL1311	cell protein precu
29	75	7.4	1642	2 T19130	hypothetical prote

ALIGNMENTS

RESULT 1

I54178
ganglioside M2 activator protein precursor - human
C:Species: Homo sapiens (man)
C:Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
C:Accession: I54178; JQ1037; S05036; S22410; S17107
R:Xie, B.; Kennedy, J.L.; McInnes, B.; Auger, D.; Mahuran, D.
Genomics 14, 796-798, 1992
A:Title: Identification of a processed pseudogene related to the functional gene encoding man chromosome 5.
A:Reference number: I54178; MUID:93052421; PMID:1427911
A:Accession: I54178
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-193 <RES>
A:Cross-references: UNIPROT:P17900; GB:L01439; NID:g183358; PIDN:AAA52767.1; PID:g183359
R:Xie, B.; McInnes, B.; Neote, K.; Lamhonwah, A.M.; Mahuran, D.
Biochem. Biophys. Res. Commun. 177, 1217-1223, 1991
A:Title: Isolation and expression of a full-length cDNA encoding the human GM2 activator
A:Reference number: JQ1037; MUID:91282768; PMID:2059210
A:Accession: JQ1037
A:Molecule type: mRNA
A:Residues: 1-18, 'A', '20-193 <XIE>
A:Cross-references: GB:M76477; NID:g183356; PIDN:AAA35907.1; PID:g183357
A:Experimental source: HeLa cell
A:Note: 19-Thr and 69-Met were also found
R:Schoeder, M.; Klina, H.; Nakano, T.; Kwon, H.; Quintern, L.E.; Gaertner, S.; Suzuki, F.
FEBS Lett. 251, 197-200, 1989
A:Title: Isolation of a cDNA encoding the human G(M2) activator protein.
A:Reference number: S05036; MUID:89325664; PMID:2753159
A:Accession: S05036
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 13-193 <SCH>
A:Cross-references: GB:X16087; NID:g31852; PIDN:CAA34215.1; PID:g31853
R:Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A:Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A:Reference number: S22410; MUID:92207171; PMID:1554364
A:Accession: S22410
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 43-142, 'VST', <NAG>
A:Cross-references: EMBL:X61094; NID:g31854; PIDN:CAA43407.1; PID:g31855
C:Comment: This protein transports GM2 ganglioside from the lysosomal membrane to hexosan
C:Genetics:
A:Gene: GDB:GM2A
A:Cross-references: GDB:120000; OMIM:272750
A:Map position: 5q32-5q33
F:1-23/Domain: signal sequence #status predicted <SIG>
F:32-193/Product: GM2 ganglioside activator protein #status predicted <GM2>
F:32-193/Product: ganglioside M2 activator #status predicted <MAT>

30 74.5 7.3 290 2 H82593 dimethyladenosine
31 74.5 7.3 388 2 F71029 hypothetical prote
32 74.5 7.3 647 2 E82579 acetyl coenzyme A
33 74.5 7.3 2482 2 I48922 cation-independent
34 74.5 7.3 2483 1 A49617 insulin-like growt
35 74.5 7.3 2823 2 F87908 protein T22A3.8 li
36 74.5 7.3 2823 2 T23084 hypothetical prote
37 74.5 7.3 3102 2 T43291 laminin alpha chai
38 74 7.3 322 2 T10179 fibrillin-like pro
39 74 7.3 432 2 A25483 env polypeptide, r
40 74 7.3 520 2 T49478 probable glutamate
41 74 7.3 785 2 A29953 alpha-1 proteinase
42 74 7.3 2187 2 T30826 nascent polypeptid
43 73.5 7.2 149 2 I69229 epidiymal secretio
44 73.5 7.2 477 1 ANRT angiotensin precur
45 73.5 7.2 515 2 AB0186 probable decarboxy

Query Match 98.7%; Score 1005; DB 2; Length 193;
 Best Local Similarity 99.0%; Pred. No. 2.7e-85;
 Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MOSLMOAPLLIALGILLATPAQAHKKPSQLSFSWDCNCFEGKDPVIRSLTLEPPIV 60
 DB 1 MOSLMOAPLLIALGILLATPAQAHKKPSQLSFSWDCNCFEGKDPVIRSLTLEPPIV 60

QY 61 PGNVTLVVVSGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIP 120
 DB 61 PGNVTLVVVSGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIP 120

QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVAVPDLPLPSWLTGNYRIESVLSGSKR 180
 DB 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVAVPDLPLPSWLTGNYRIESVLSGSKR 180

QY 181 LGCIKIAASLKI 193
 DB 181 LGCIKIAASLKI 193

RESULT 2
 S22411
 ganglioside M2 activator protein (clone pGAP2) - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
 C:Accession: S22411; S17108
 R:Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
 Biochem. J. 282, 807-813, 1992
 A:Title: Evidence for two cDNA clones encoding human GM2-activator protein.
 A:Reference number: S22410; MUID:92207171; PMID:1554364
 A:Accession: S22411
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-200 <NAG>
 A:Cross-references: UNIPROT:PI7900; EMBL:X61095; NID:g31956; PIDN:CAA43408.1; PID:g31957

Query Match 97.8%; Score 996; DB 2; Length 200;
 Best Local Similarity 97.4%; Pred. No. 1.9e-84;
 Matches 188; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MOSLMOAPLLIALGILLATPAQAHKKPSQLSFSWDCNCFEGKDPVIRSLTLEPPIV 60
 DB 8 MOSLMOAPLLIALGILLATPAQAHKKPSQLSFSWDCNCFEGKDPVIRSLTLEPPIV 67

QY 61 PGNVTLVVVSGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIP 120
 DB 68 PGNVTLVVVSGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIP 127

QY 121 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVAVPDLPLPSWLTGNYRIESVLSGSKR 180
 DB 128 TGEPCPEPLRTYGLPCHCPKFGTYSLPKSEFVAVPDLPLPSWLTGNYRIESVLSGSKR 187

QY 181 LGCIKIAASLKI 193
 DB 188 LGCIKIAASLKI 200

RESULT 3
 S13195
 ganglioside M2 activator protein - human
 C:Species: Homo sapiens (man)
 C:Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
 C:Accession: S13195
 R:Fuerst, W.; Schubert, J.; Machleidt, W.; Meyer, H.E.; Sandhoff, K.
 Eur. J. Biochem. 192, 709-714, 1990
 A:Title: The complete amino-acid sequences of human ganglioside GM2 activator protein and
 A:Reference number: S13195; MUID:91006185; PMID:2209618
 A:Accession: S13195
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-162 <FUE>

A:Cross-references: UNIPROT:PI7900

Query Match 82.9%; Score 844; DB 2; Length 162;
 Best Local Similarity 98.1%; Pred. No. 1.5e-70;
 Matches 157; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 32 SFSFWDNCFEGKDPVIRSLTLEPPIVPGNVTLVSVGTSVPLSSPLKVDLVLEKEVA 91
 DB 1 SFSFWDNCFEGKDPVIRSLTLEPPIVPGNVTLVSVGTSVPLSSPLKVDLVLEKEVA 60

QY 92 GLWIKIPCTDIYIGSCTFEHFCVDLMDLIPTBPCPEPLRTYGLPCHCPKFGTYSLPKSE 151
 DB 61 GLWIKIPCTDIYIGSCTFEHFCVDLMDLIPTBPCPEPLRTYGLPCHCPKFGTYSLPKSE 120

QY 152 FAVPDLPLPSWLTGNYRIESVLSGSKRLGCIKIAASLK 191
 DB 121 FAVPDLPLPSWLTGNYRIESVLSGSKRLGCIKIAASLK 160

RESULT 4

S35613
 ganglioside M2 activator protein precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 07-May-1999
 C:Accession: S35613
 R:Belachova, G.; Stirling, J.L.; Orlicchio, A.; Beccari, T.
 Biochem. J. 294, 227-230, 1993
 A:Title: Cloning and sequence analysis of a cDNA clone coding for the mouse G(M2) activa
 A:Reference number: S35613; MUID:93371367; PMID:7689829
 A:Accession: S35613
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-193 <BEL>

Query Match 72.5%; Score 738.5; DB 2; Length 193;
 Best Local Similarity 71.8%; Pred. No. 9.8e-61;
 Matches 135; Conservative 23; Mismatches 27; Indels 3; Gaps 1;

QY 8 PLLIALGLLA---TPQAHLKKPSQLSFSWDCNCFEGKDPVIRSLTLEPPIVPGNV 64
 DB 5 PLLIALGLLA---TPQAHLKKPSQLSFSWDCNCFEGKDPVIRSLTLEPPIVPGNV 64

QY 65 TSVSVGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIP 124
 DB 65 VSVLEKTSVPLTAPOKVELTVEKEVAGLWIKIPCTDIYIGSCTFEHFCVDLMDLIP 124

QY 125 CPEPLRTYGLPCHCPKFGTYSLPKSEFVAVPDLPLPSWLTGNYRIESVLSGSKRLGCI 184
 DB 125 CPEPLRTYGLPCHCPKFGTYSLPKSEFVAVPDLPLPSWLTGNYRIESVLSGSKRLGCI 184

QY 185 KIAASLKG 192
 DB 185 KIAASLKG 192

RESULT 5

T49126
 hypothetical protein F26G5.50 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
 C:Accession: T49126
 R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z25017
 A:Accession: T49126
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-152 <DAN>
 A:Cross-references: UNIPROT:Q9LXQ2; EMBL:AL353814; GSPDB:GN00061; ATSP:F26G5.50
 A:Experimental source: cultivar Columbia; EAC clone F26G5
 C:Genetics:
 A:Gene: ATSP:F26G5.50
 A:Map position: 3

A; Introns: 31/1; 64/1; 117/3

Query Match 11.1%; Score 113; DB 2; Length 152;
Best Local Similarity 24.5%; Pred. No. 0.0033;
Matches 45; Conservative 25; Mismatches 64; Indels 50; Gaps 7;
QY 3 SLMOA-PLIALGLLLATPAQAHLKKPSQLSSFSWDCNCFEGKDPVIRSLTLEPPIVVP 61
DB 2 AISQAQPLLLLLLVFLPA-----LHATSFYCDKRLDPVKVGVKISPDVWSG 52
QY 62 GNVLSVVGSTSVPLSSPKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDVLMILPT 121
DB 53 AAATEKIFGSGEDISGGKVVIRVL-----YVG-----IPVHTET 87
QY 122 GPCPEPLRTYGLPCHCPKEGTYSLPKSEFAVPDLEPLSWLTITGNRYLESVLS-SSGKR 180
DB 88 HDLCDETA-----CPVAPGSFVLHSQ-----TLFSITPPGTYTLKMTINDKNGR 133
QY 181 LGCI 184
DB 134 LTCI 137

RESULT 6

G72563
hypothetical protein APE1793 - Aeropyrum pernix (strain K1)
C; Species: Aeropyrum pernix
C; Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
R; Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, Y.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A; Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A; Reference number: A72450; MUID: 99310339; PMID: 10382966
A; Accession: G72563
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-171 <KAW>
A; Cross-references: UNIPROT:Q9YB02; DDBJ:AP0000062; NID:G5105244; PIDN:BAAR0796.1; PID:dl
A; Experimental source: strain K1
C; Genetics:
A; Gene: APE1793

Query Match 9.2%; Score 94; DB 2; Length 171;
Best Local Similarity 25.8%; Pred. No. 0.22;
Matches 32; Conservative 17; Mismatches 43; Indels 32; Gaps 4;
QY 10 LIALGLLLATPAQAHLKKPSQLSSFS-----WDCNCFEGKDPVIRSLT 52
DB 3 MLCSLILETSTLTSTSPISLSSTFTRAGRFLAPLILASVTWGNWSCFSGLKPLMLVRLS 62
QY 53 L-----EPPPIVPGNVT--LSVVGSTSVPLSSPKVDLVLEKEVAGLWIKIPCTDYI 103
DB 63 TSSILRSSPAPPPTPTHTTGLSLGKAPSLSLNLSLIDAN-----RPFISAAT 116
QY 104 GSCT 107
DB 117 GSLT 120

RESULT 7

T34048
hypothetical protein W08F4.6 - Caenorhabditis elegans
C; Species: Caenorhabditis elegans
C; Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C; Accession: T34048
R; Madson, C.; Graves, T.
submitted to the EMBL Data Library, April 1997
A; Description: The sequence of C. elegans cosmid W08F4.
A; Reference number: Z21468
A; Accession: T34048
A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: DNA

A; Residues: 1-294 <MAD>

A; Cross-references: UNIPROT:O02106; EMBL:AF000266; PIDN:AAAC71171.1; GSPDB:GN00020; CESP:V
A; Experimental source: strain Bristol N2; clone W08F4

C; Genetics:

A; Gene: W08F4.6

A; Map position: 2

A; Introns: 58/3; 138/2; 229/3

C; Superfamily: Caenorhabditis elegans hypothetical protein W08F4.6

Query Match 9.0%; Score 92; DB 2; Length 294;
Best Local Similarity 21.9%; Pred. No. 0.64;
Matches 59; Conservative 27; Mismatches 68; Indels 116; Gaps 14;

QY 9 LIALGLLLATPAQAHLKKPSQLSSFSWDCNCFEGKDP-----AVIR----- 49
DB 6 LLLAVGTVV---AQNPLTK-----ANNEAVPGVQPFWEKYQTGHHGVIRGWFQRCA 55
QY 50 -----SLTLEPPIVVPG-----NVTLSVVGSTSVPLSSPKVDLVLE----- 87
DB 56 SEQWTVVNVVSNIVIPDPYPRFGPIFENVMTDV--SEDLPLD--KVEMDLEVRHVT 110
QY 88 -KEVAGLWIKIP-----TDYIGSCT-----FEHFCVDVL 115
DB 111 NKQSGKGMQVIPCQGWNIIDGCGVGSNNVWVWITAQNVSTNLVKNAEMLQIYSRCDML 170
QY 116 DM-----LIPTGCPCEPLRTYGLPCHCPKEGTYSLPKSEFAVPD 156
DB 171 DKNEAVGAHKYVKDKRKANDFLKQNLCPFPKGHWTMTFSKVF--SSDLPSKFFG-PL 227
QY 157 LELPSWLTITGNRYLESVLSSSGKRLGCIKI 186
DB 228 QSEYWLTF-----SFTDGKDKKLGARL 251

RESULT 8

AI0201
beta-galactosidase (EC 3.2.1.23) [imported] - Yersinia pestis (strain CO92)
C; Species: Yersinia pestis
C; Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C; Accession: AI0201
R; Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; E
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, E
Nature 413, 523-527, 2001
A; Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A; Reference number: AB00001; MUID: 21470413; PMID: 11586360
A; Accession: AI0201
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-1060 <KUR>
A; Cross-references: UNIPROT:Q8ZFP0; GB:AL590842; PIDN:CAC90476.1; PID:gl5979691; GSPDB:G

C; Genetics:

A; Gene: lacZ

C; Superfamily: beta-galactosidase

C; Keywords: glycosidase; hydrolase

Query Match 8.5%; Score 86.5; DB 2; Length 1060;
Best Local Similarity 24.2%; Pred. No. 9.6;
Matches 43; Conservative 18; Mismatches 54; Indels 63; Gaps 10;

QY 20 PAQAHLKKPSQLSSFSWDCNCFEGKDPVIRSLTLEPPIVPGNVTLVVGSTSVPLSSP 79
DB 8 PLQVQLSLPQLISREDWEN-----PQITQYHRLAHP---PFH-SWRDVESAKDRSP 57
QY 80 LKVDLVLEKEVAGLWIKIPCTDYIGSCTFEH-----CDVLDMLIPTGCPCE 127
DB 58 -----QQQTINGLW-----SFSYFTQPEAVPEHWRCDLAE-----AKPLPV 94
QY 128 P-----LRTYGLPCHCPKEGTYSLPKSEFAVPDLE-----LPSWLTITGNRI 170
DB 95 PANWQLHGYPDIYTNIQ---YPIVNPVRPVDLNPCTGCYSRDLTLEFSLWLASGKTRI 149

RESULT 9

THHUB
 A:Title: thrombomodulin precursor [validated] - human
 C:Species: Homo sapiens (man)
 C>Date: 31-Dec-1998 #sequence revision 12-May-1995 #text change 09-Jul-2004
 C:Accession: A41442; A28307; A29680; A27073; JX0264; S38954
 R:Shirai, T.; Shiojiri, S.; Ito, H.; Yamamoto, S.; Kusumoto, H.; Deyashiki, Y.; Maruyama
 J. Biochem. 103, 281-285, 1988
 A:Title: Gene structure of human thrombomodulin, a cofactor for thrombin-catalyzed activ
 A:Reference number: A41442; MUID:88227901; PMID:2836377
 A:Accession: A41442
 A:Molecule type: DNA
 A:Residues: 1-575 <SHI>
 A:Cross-references: UNIPROT:P07204; DBJ:D00210; NID:g220126; PIDN:BAA00149.1; PID:g2201
 R:Jackman, R.W.; Beeler, D.L.; Fritze, L.; Soff, G.; Rosenberg, R.D.
 Proc. Natl. Acad. Sci. U.S.A. 84, 6425-6429, 1987
 A:Title: Human thrombomodulin gene is intron depleted; nucleic acid sequences of the cDN
 A:Reference number: A28307; MUID:87317665; PMID:2819876
 A:Accession: A28307
 A:Molecule type: mRNA
 A:Residues: 1-472, 'A', 474-575 <JAC>
 A:Cross-references: GB:J02973; NID:g339658; PIDN:AAA61175.1; PID:g339659
 R:Suzuki, K.; Kusumoto, H.; Deyashiki, Y.; Nishioka, J.; Maruyama, I.; Zushi, M.; Kawaba
 EMBO J. 6, 1891-1897, 1987
 A:Title: Structure and expression of human thrombomodulin, a thrombin receptor on endoth
 A:Reference number: A29680; MUID:88004395; PMID:2820710
 A:Accession: A29680
 A:Molecule type: mRNA
 A:Residues: 1-575 <SUZ>
 A:Cross-references: GB:X05495; NID:g37123; PIDN:CAA29045.1; PID:g736251
 A:Experimental source: lung endothelium
 A:Note: part of this sequence, including the amino end of the mature protein, were deter
 R:Wen, D.; Dittman, W.A.; Ye, R.D.; Deaven, L.L.; Majerus, P.W.; Sadler, J.E.
 Biochemistry 26, 4350-4357, 1987
 A:Title: Human thrombomodulin: complete cDNA sequence and chromosome localization of the
 A:Reference number: A27073; MUID:88024950; PMID:2822087
 A:Accession: A27073
 A:Molecule type: mRNA
 A:Residues: 1-472, 'A', 474-575 <WEN>
 A:Cross-references: GB:M16552; NID:g339656; PIDN:AAB59508.1; PID:g339657
 A:Experimental source: placenta
 A:Note: parts of this sequence were determined by protein sequencing
 R:Yamamoto, S.; Mizoguchi, T.; Tamaki, T.; Ohkuchi, M.; Kimura, S.; Aoki, N.
 J. Biochem. 113, 433-440, 1993
 A:Title: Urinary thrombomodulin, its isolation and characterization.
 A:Reference number: JX0264; MUID:93293792; PMID:8390446
 A:Accession: JX0264
 A:Molecule type: protein; mRNA
 A:Residues: 19-472, 'A', 474-486 <YAM>
 A:Experimental source: urine
 A:Note: the urinary form appears to be identical with that circulating in plasma
 R:Gerlitz, B.; Hassell, T.; Vlahos, C.J.; Parkinson, J.F.; Bang, N.U.; Grinnell, B.W.
 Biochem. J. 295, 131-140, 1993
 A:Title: Identification of the predominant glycosaminoglycan-attachment site in soluble
 serine.
 A:Reference number: S38954; MUID:94023900; PMID:8216207
 A:Accession: S38954
 A:Molecule type: protein
 A:Residues: 475-491, 'X', 493-494 <GER>
 A:Note: the residue designated 'X' was determined to be a Ser with covalently bound chor
 R:Meininger, D.P.; Komives, E.A.
 submitted to the Brookhaven Protein Data Bank, September 1995
 A:Reference number: A67369; PDB:1ZQA
 R:Tulinsky, A.; Mathews, I.I.
 submitted to the Brookhaven Protein Data Bank, August 1994
 A:Reference number: A52804; PDB:1HLT
 R:Hrabal, R.; Komives, E.A.; Ni, F.
 submitted to the Brookhaven Protein Data Bank, November 1995
 A:Reference number: A65583; PDB:1FGD
 A:Contents: annotation; conformation by (1)H-NMR, residues 427-444
 R:Hrabal, R.; Komives, E.A.; Ni, F.

Protein Sci. 5, 195-203, 1996
 A:Title: Structural resiliency of an EGF-like subdomain bound to its target protein, th
 A:Reference number: A58595; MUID:96276211; PMID:8745396
 A:Contents: annotation; conformation by (1)H-NMR
 C:Genetics:
 A:Gene: GDB:THBD
 A:Cross-references: GDB:119613; OMIM:188040
 A:Map position: 20p11.2-20p11.2
 A:Introns: #status absent
 C:Complex: homodimer, urinary form
 C:Function:
 A:Description: inhibits thrombin activation of fibrinogen; cofactor for thrombin activa
 A:Pathway: blood coagulation moderation
 A:Note: the membrane-bound form is located on the endothelium luminal surface of arteri
 C:Superfamily: thrombomodulin; C-type lectin homology; EGF homology
 C:Keywords: anticoagulant; beta-hydroxyasparagine; beta-hydroxyaspartic acid; blood coag
 e protein
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-575/Product: thrombomodulin, membrane-bound form #status predicted <MAT>
 F:19-513/Domain: extracellular #status predicted <EXT>
 F:19-486/Product: thrombomodulin, urinary form #status experimental <MAU>
 F:24-167/Domain: C-type lectin homology <LCH>
 F:177-199/Region: PEST sequence
 F:201-233/Region: PEST sequence
 F:245-280/Domain: EGF homology <EG1>
 F:288-323/Domain: EGF homology <EG2>
 F:329-362/Domain: EGF homology <EG3>
 F:369-404/Domain: EGF homology <EG4>
 F:408-439/Domain: EGF homology <EG5>
 F:445-480/Domain: EGF homology <EG6>
 F:485-513/Region: PEST sequence
 F:517-539/Domain: transmembrane #status predicted <TMN>
 F:540-575/Domain: intracellular #status predicted <INT>
 F:47,115,116,382,409/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:174,225,411,504/Binding site: carbohydrate (Thr) (covalent) #status predicted
 F:245-256,252-265,267-280,288-296,292-308,310-323,329-340,336-349,351-362,369-378,374-38
 F:334,498/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:342/Modified site: erythro-beta-hydroxyasparagine (Asn) #status experimental
 F:490,499/Binding site: chondroitin sulfate (Ser) (covalent) (partial) #status experimen
 Query Match 8.1%; Score 82; DB 1; Length 575;
 Best Local Similarity 23.3%; Pred. No. 12;
 Matches 49; Conservative 23; Mismatches 74; Indels 64; Gaps 12;
 QY 8 PLLIALGELLAT-----PAQAHLLKPSQLSFSWNCPEGKDPVAVIRSLTLEPDPITV 59
 DB 135 PLCAVAVSAAEATVPSEPIWEEQCEVADGFLCEFFH-----PATCRPLAVEPGAAA 186
 QY 60 VPGNVTLVS-----VGSTSVPLSSPLKVDIVL-----EKEVAGLWI 95
 DB 187 AAVSITVGTTPAARGADFOALPVGSSAA--VAPLGLQLMCTAPPAGVGVGHWRAPGAW- 243
 QY 96 KIPCTDIYGSTCFRHFCDVLDMLIPTGEPCEP----LRTYGLPCHPCKFKGTYSLPKSE 151
 DB 244 --DCSVENGCC--EHAACNA---IPGAPRCQCPAGAAALQADGRSCTASATSCNDLCE-H 294
 QY 152 FAVPDLEPLPSWLTITGNVRIESVLSKRL 181
 DB 295 FCVPNPDP-----GSY---SCMCETGYRL 316
 RESULT 10
 A47651
 zinc-finger protein - black rat
 C:Species: Rattus rattus (black rat, roof rat)
 C>Date: 02-Jul-1996 #sequence revision 02-Jul-1996 #text change 09-Jul-2004
 C:Accession: A47651
 R:Lipkin, S.M.; Naar, A.M.; Kalla, K.A.; Sack, R.A.; Rosenfeld, M.G.
 Genes Dev. 7, 1674-1687, 1993
 A:Title: Identification of a novel zinc finger protein binding a conserved element critic
 A:Reference number: A47651; MUID:93380644; PMID:8370519
 A:Accession: A47651

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2180 <RES>
A;Cross-references: UNIPROT:Q63753; GB:L23077; NID:9949987; PIDN:AAA74461.1; PID:g349558

Query Match 8.0%; Score 81; DB 2; Length 2180;
Best Local Similarity 24.1%; Pred. No. 74;
Matches 41; Conservative 30; Mismatches 61; Indels 38; Gaps 9;

QY 2 QSLMOA---PLLTALGILLATPAQAHLKPSQSSFWNDNCFEGKDPVAVIRSLTLRPDPI 58
DB 871 QELLQTNQPSLLA--SMHITTTVNMQQPQQ--SNCPETCF--KDPFQLLSVNRPN 925
QY 59 VFGNVTLVVGSTSVPLSLPLKVDLVLEKEVAGLMIKIPCTDIYIGCTFEHFCVDLMDL 118
DB 926 FLSTPRCDVSNFAGVSQEGSEIIKALETAGI-----PSTFES-AEMLSQV 973
QY 119 IPTG-----EPCPELRTYGLP--CHCFKFGCTYSLPKSEFAVPD 156
DB 974 VPIGSVSDAAQVNAEGMPGPTVTPLLQTVCH-----TNTSPSSQNQTPN 1017

RESULT 11
AD3629
vdcC protein [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C;Accession: AD3629
R;DeiVecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, M.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, P.; Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A;Reference number: AD3252; PMID:11756688
A;Accession: AD3629
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-504 <KUR>
A;Cross-references: UNIPROT:Q8YBE3; GB:AE008918; PIDN:AAL54199.1; PID:g17985167; GSPDB:G000000000
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0957
A;Map position: 11
C;Superfamily: conserved hypothetical protein sll0936

Query Match 7.9%; Score 80.5; DB 2; Length 504;
Best Local Similarity 27.8%; Pred. No. 14;
Matches 20; Conservative 16; Mismatches 33; Indels 3; Gaps 1;

QY 8 PLLTALGILLATPAQAHLKPSQSSFWNDNCFEGKDPVAVIRSLTLRPDPIVPGNVTL 67
DB 224 PVAIAIGVDPATILAAVMPLEPMSELAFLGGLGRPCVTQGRV---PLMVPANAEIV 280
QY 68 VVGSTSVPLSSP 79
DB 281 LEGRVSAQTQAP 292

RESULT 12
T49440
alcohol dehydrogenase (EC 1.1.1.1) I adh1 B17C10.210 [similarity] - Neurospora crassa
N;Alternate names: protein B17C10.210
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C;Accession: T49440
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49440
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-353 <SCH>
A;Cross-references: UNIPROT:Q9P6C8; EMBL:AL355926; GSPDB:GN00116; NCSP:B17C10.210
A;Experimental source: BAC clone B17C10; strain OR74A

C;Genetics:
A;Gene: NCSP:B17C10.210
A;Map position: 6
A;Introns: 21/1
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: NAD; oxidoreductase; zinc
F;47,70,157/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 7.8%; Score 79; DB 2; Length 353;
Best Local Similarity 26.8%; Pred. No. 13;
Matches 52; Conservative 26; Mismatches 62; Indels 54; Gaps 16;

QY 36 WDNCFEGK-DPAVIRSLTLE---PDPVIVPGNVTLVVGSTSV-----PLSS--PL-- 80
DB 11 WAQVVEKGGPVVFKQIPVOKPGPDEVLI--NVKYSVCHTDLHAMKGDWFLATKMLV 68
QY 81 -----KVDLVLEKEV---AGL-WIKIPCTDIYIGCTFEHFCVDLMDLPTGP-C 125
DB 69 GHEGAGVWVAKGELVTEVEGDHAGIKWLN-----GSLACSFQAD-----EPLC 115
QY 126 PEPRLTYGLPCHCFKFGCTYSLPKSEFA--VP---DLELPSWLTGNYRIESVLSGSKR 180
DB 116 PHALLS-GYTVDGSGFQQ--YAIKAHVAKIPKGCDDLETTAPVLCAGITVYKLGKESGVR 172
QY 181 LG-CIKIAASLKI 193
DB 173 PGQCVAVIAGGGGL 186

RESULT 13
C71402
probable glucosyltransferase - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: Columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 09-Jul-2004
C;Accession: C71402
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks, P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giele, avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B. Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, erthof, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ansc C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A;Reference number: A71400; PMID:98121113; PMID:9461215
A;Accession: C71402
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-456 <BEV>
A;Cross-references: UNIPROT:Q23270; GB:Z97335; NID:g2244747; PID:g2244766
C;Genetics:
A;Map position: 4COP9-4G3845
C;Superfamily: flavonol O3-glucosyltransferase

Query Match 7.8%; Score 79; DB 2; Length 456;
Best Local Similarity 21.9%; Pred. No. 17;
Matches 50; Conservative 23; Mismatches 65; Indels 90; Gaps 10;

QY 15 LLLATPAQAHLKPSQLS-----SFSW-----D 37
DB 15 LLVTFPAQGHINPALQANRLIHGATVTVSTAVSAHRMGEPSTKGLSFAWTDGDD 74
QY 38 NCFEGKDPVAVIRSLTLE-----LTLEPPIVPGNVTLVVGSTSVPLSS 78
DB 75 GLKSPEDQKIYMSSELKRCGSNALRDIKANLDAATETEPIT-----GVISVLVPMWS 127
QY 79 PLKVDLVLEKEVAGLMIKIPCT---DYIGCTFEHFCVDLMDLPTGPCEPLRTY 133
DB 128 TVAREFHLPITL--LWIE-PATVLDIYYVNTSTKHLFDV-----EPIKLPK 172
QY 134 LPCHCFKFGCTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKRL 181
DB 173 LPLIT-----TGDLP--SFLQSPKALPSALVTLREHIALETESNPKI 213

1

db 483 HCCPSLTLTAPSGNGTESATPLGYPYLTNYPCDANKPIPSQSFSTYA-----FCDPDT 537

db 483 HCCPSLTLTAPSGNGTESATPLGYPYLTNYPCDANKPIPSQSFSTYA-----FCDPDT 537

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 05:42:53 ; Search time 29.6589 Seconds
(without alignments)
3744.152 Million cell updates/sec

Title: US-10-030-937-9

Perfect score: 1018

Sequence: 1 MQSLMQAPLIIALGLLIATP.....LSSSGKRLGCIKIAASLKGI 193

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1005	98.7	193	1	SAP3 HUMAN
2	987	97.0	189	2	Q6LBI5
3	987	97.0	189	2	Caa43994
4	907.5	89.1	190	2	Q8HXX6
5	739.5	72.6	199	2	Q6IN37
6	739.5	72.6	199	2	Q8CJH4
7	737.5	72.4	193	1	SAP3 MOUSE
8	602	59.1	151	2	Q6Q7X5
9	602	59.1	151	2	AA564350
10	544	53.4	103	2	Q14427
11	519	51.0	146	2	Q6Q7X4
12	519	51.0	146	2	AA564351
13	503.5	49.5	197	2	Q6GLN6
14	257	25.2	217	2	Q75R48
15	257	25.2	217	2	BAD16601
16	222	21.8	47	2	Q8I028
17	138	13.6	40	2	Q29283
18	117	11.5	242	2	Q6J1S7
19	117	11.5	242	2	AAT35232
20	113	11.1	152	2	Q9LXQ2
21	111	10.9	26	2	Q8I029
22	107.5	10.6	155	2	Q8I6V6
23	105	10.3	480	2	Q8VQNI
24	103.5	10.2	243	2	Q7S5Y5
25	103.5	10.2	331	2	Q6NE93
26	103.5	10.2	331	2	CAF06269
27	102	10.0	480	2	Q700B8
28	102	10.0	480	2	Q9WXD3
29	102	10.0	480	2	CAF74787
30	94	9.2	171	2	Q9YB02
31	93	9.1	556	2	O87548

RESULT 1

SAP3 HUMAN

ID SAP3 HUMAN STANDARD; PRT; 193 AA.
AC P17900; Q14426; Q14428;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate activator protein) (Shingolipid activator protein 3) (SAP-3).
GN Name=GM2A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RX MEDLINE=91282768; PubMed=2059210;
RA Xie B., McInnes B., Neote K., Lambonwah A.-M., Mahuran D.;
RT "Isolation and expression of a full-length cDNA encoding the human G-M2 activator protein."
RL Biochem. Biophys. Res. Commun. 177:1217-1223(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K., Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the human GM2 activator protein."
RL FEBS Lett. 289:260-264(1991).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RC TISSUE=Placenta;
RX MEDLINE=92207171; PubMed=1554364;
RA Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
RT "Evidence for two cDNAs encoding human GM2-activator protein."
RL Biochem. J. 282:807-813(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93052421; PubMed=1427911;
RA Xie B., Kennedy J.L., McInnes B., Auger D., Mahuran D.J.;
RT "Identification of a processed pseudogene related to the functional gene encoding the GM2 activator protein: localization of the pseudogene to human chromosome 3 and the functional gene to human chromosome 5."
RL Genomics 14:796-798(1992).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=99294584; PubMed=10364519;
RA Chen B., Rigat B., Curry C., Mahuran D.J.;
RT "Structure of the GM2A gene: identification of an exon 2 nonsense mutation and a naturally occurring transcript with an in-frame deletion of exon 2."
RL Am. J. Hum. Genet. 65:77-87(1999).
RN [6]

32 92 294 2 002106
33 92 373 2 Q8JVA9
34 91.5 9.0 650 2 Q6ND98
35 91.5 9.0 650 2 CAE25655
36 88 339 2 Q8QX8
37 88 8.6 850 2 Q6F2P9
38 87.5 8.5 323 2 Q8EL09
39 86.5 8.5 723 2 Q960N8
40 86.5 8.5 768 2 Q46043
41 86.5 8.5 1060 2 Q8ZFP0
42 86.5 8.5 1060 2 AAS62012
43 86 8.4 143 1 ALL2_PSOOV
44 86 8.4 1403 2 Q922D3
45 86 8.4 1898 2 Q8CHF2

ALIGNMENTS

RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McBean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[7]
RP SEQUENCE OF 15-193 FROM N.A.
RX MEDLINE=89325664; PubMed=2753159;
RA Schroeder M., Klina H., Nakano T., Kwon H., Quintern L.E.,
RA Gaertner S., Suzuki K., Sandhoff K.;
RT "Isolation of a cDNA encoding the human GM2 activator protein.";
RL FEBS Lett. 251:197-200(1989).
[8]
RP SEQUENCE OF 32-193.
RC TISSUE=Kidney;
RX MEDLINE=91006185; PubMed=2209618;
RA Furst W., Schubert J., Machleidt W., Meyer H.E., Sandhoff K.;
RT "The complete amino-acid sequences of human ganglioside GM2 activator
RL protein and cerebroside sulfate activator protein.";
RN Eur. J. Biochem. 192:709-714(1990).
[9]
RP X-RAY CRYSTALLOGRAPHY (2.0 ÅNGSTROMS).
RX MEDLINE=20545600; PubMed=11090283; DOI=10.1006/jmbi.2000.4225;
RA Wright C.S., Li S.-C., Rastinejad F.;
RT "Crystal structure of human GM2-activator protein with a novel beta-
RL cup topology.";
RN J. Mol. Biol. 304:411-422(2000).
[10]
RP VARIANT TSD-AB ARG-138.
RX MEDLINE=92008638; PubMed=1915858;
RA Schroeder M., Schnabel D., Suzuki K., Sandhoff K.;
RT "A mutation in the gene of a glycolipid-binding protein (GM2
RL activator) that causes GM2-gangliosidosis variant AB.";
RN FEBS Lett. 290:11-13(1991).
[11]
RP VARIANT TSD-AB PRO-169.
RX MEDLINE=94063850; PubMed=8244332;
RA Schroeder M., Schnabel D., Hurwitz R., Young E., Suzuki K.,
RA Sandhoff K.;
RT "Molecular genetics of GM2-gangliosidosis AB variant: a novel mutation
RL and expression in BHK cells.";
RN Hum. Genet. 92:437-440(1993).
[12]
RP VARIANT TSD-AB LYS-88 DEL.
RX MEDLINE=97055887; PubMed=8900233;
RA Schepers U., Glombitza G., Lemm T., Hoffmann A., Chabas A., Orand P.,
RA Sandhoff K.;
RT "Molecular analysis of a GM2-activator deficiency in two patients with
RL GM2-gangliosidosis AB variant.";
RN Am. J. Hum. Genet. 59:1048-1056(1996).
CC -!- FUNCTION: Binds gangliosides and stimulates ganglioside GM2
CC degradation. It stimulates only the breakdown of ganglioside GM2
CC and glycolipid GA2 by beta-hexosaminidase A. It extracts single
CC GM2 molecules from membranes and presents them in soluble form to
CC beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and

conversion to GM3.
-!- SUBCELLULAR LOCATION: Lysosomal.
-!- DISEASE: Defects in GM2A are the cause of Tay-Sachs disease AB
variant (TSD-AB) [MIM:272750]; also known as GM2-gangliosidosis
type AB.
-!- DATABASES: NAME=GM2Adb; NOTE=GM2A mutation database;
WWW="http://www.hexdb.mcgill.ca/?Topic=GM2Adb&Page=MutationSubmission".

This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; M76477; AAA35907.1; -;
DR EMBL; X62078; CAA43993.1; -;
DR EMBL; X61095; CAA43408.1; ALT_INIT.
DR EMBL; L01439; AAA52767.1; -;
DR EMBL; AF124719; AAD25741.1; -;
DR EMBL; AF124717; AAD25741.1; JOINED.
DR EMBL; AF124718; AAD25741.1; JOINED.
DR EMBL; BC009273; AAH09273.1; -;
DR EMBL; X16087; CAA34215.1; -;
DR PIR; I54178; I54178.
DR PIR; S13195; S13195.
DR PIR; S22411; S22411.
DR PDB; 1G13; X-ray; A/B/C=32-193.
DR Genew; HGNC:4367; GM2A.
DR MIM; 272750; -;
CO; GO:0005764; C:lysosome; NAS.
DR GO; GO:0030290; F:phospholipid activator protein activity; NAS.
DR GO; GO:0019377; P:glycolipid catabolism; NAS.
DR GO; GO:0030149; P:phospholipid catabolism; NAS.
DR InterPro; IPR003172; E1_DerP2_DerF2.
DR SMART; SM00737; MC; 1.
KW 3D-structure; Direct protein sequencing; Disease mutation;
KW Glycoprotein; GM2-gangliosidosis; Lysosome; Polymorphism; Signal;
KW Sphingolipid metabolism.
FT CHAIN 1 31 Ganglioside GM2 activator.
FT SIGNAL 32 193
FT DISULFID 39 183
FT DISULFID 99 106
FT DISULFID 112 138
FT DISULFID 125 136
FT CARBOHYD 63 63
FT VARIANT 19 19 N-linked (GlcNAc...).
FT T -> A.
FT VARIAT 32 33 /FTId=VAR_013830.
FT Missing (in 80% of the protein).
FT VARIAT 88 88 /FTId=VAR_006946.
FT Missing (in TSD-AB).
FT VARIAT 138 138 /FTId=VAR_011697.
FT Missing (in TSD-AB).
FT VARIAT 169 169 /FTId=VAR_006947.
FT R -> P (in TSD-AB).
FT CONFLICT 59 59 /FTId=VAR_011698.
FT CONFLICT 69 69 V -> I (in Ref. 3).
FT STRAND 35 38 V -> M (in Ref. 3).
FT TURN 41 43
FT STRAND 46 54
FT STRAND 58 59
FT STRAND 63 72
FT STRAND 76 76
FT STRAND 81 90
FT TURN 91 92
FT STRAND 93 96
FT TURN 100 100
FT STRAND 101 102
FT STRAND 103 103
FT STRAND 107 108
FT TURN 109 110


```

ET  HELIX      111      118
FT  TURN      121      122

Query Match      98.7%; Score 1005; DB 1; Length 193;
Best Local Similarity 99.0%; Pred. No. 1.7e-83;
Matches 191; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1  MQSLMQAPLLIALGILLATPAQAHLKPSQSSFSWDCNCFEGKDPAVIRSLTLEPDPITW 60
DB  1  MQSLMQAPLLIALGILLATPAQAHLKPSQSSFSWDCNCFEGKDPAVIRSLTLEPDPITW 60

QY  61  PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGCTFEHFCVDLMDLIP 120
DB  61  PGNVTLVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGCTFEHFCVDLMDLIP 120

QY  121  TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180
DB  121  TGEPCPEPLRTYGLPCHCPFKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKR 180

QY  181  LGCIIKIAASLKGI 193
DB  181  LGCIIKIAASLKGI 193

RESULT 2
Q6LBL5  PRELIMINARY; PRT; 189 AA.
AC  Q6LBL5;
DT  05-JUL-2004 (TrEMBLrel. 27, Created)
DT  05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT  05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE  GM2 activator protein.
OS  Homo sapiens (human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=92008637; PubMed=1915857;
RA  Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA  Sandhoff K.;
RT  "Characterization of full-length cDNAs and the gene coding for the
RT  human GM2 activator protein."
RL  FEBS Lett. 289:260-264(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93277527; PubMed=8503891;
RA  Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA  Sandhoff K.;
RT  "Over-expression of a functionally active human GM2-activator protein
RT  in escherichia coli."
RL  Biochem. J. 292:571-576(1993).
DR  EMBL; X62078; CAA43994.1; -.
DR  SMART; SM00737; ML; 1.
SQ  SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match      97.0%; Score 987; DB 2; Length 189;
Best Local Similarity 98.9%; Pred. No. 7.1e-82;
Matches 187; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  5  MQAPLLIALGILLATPAQAHLKPSQSSFSWDCNCFEGKDPAVIRSLTLEPDPIVPGNV 64
DB  1  MQAPLLIALGILLATPAQAHLKPSQSSFSWDCNCFEGKDPAVIRSLTLEPDPIVPGNV 60

QY  65  TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGCTFEHFCVDLMDLPTGEP 124
DB  61  TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGCTFEHFCVDLMDLPTGEP 120

QY  125  CPEPLRTYGLPCHCPFKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKRLGCI 184
DB  121  CPEPLRTYGLPCHCPFKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKRLGCI 180

QY  185  KIAASLKGI 193
DB  181  KIAASLKGI 189

RESULT 4
Q8HXX6  PRELIMINARY; PRT; 190 AA.
AC  Q8HXX6;
DT  01-MAR-2003 (TrEMBLrel. 23, Created)
DT  01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Ganglioside GM2 activator.
GN  Name=gm2a;
OS  Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC  Cercopitheciinae; Macaca.
OX  NCBI_TaxID=9541;
RN  [1]
RP  SEQUENCE FROM N.A.

```

```

QY  185  KIAASLKGI 193
DB  181  KIAASLKGI 189

RESULT 3
CAA43994  PRELIMINARY; PRT; 189 AA.
AC  CAA43994;
DT  02-MAR-2004 (TrEMBLrel. 27, Created)
DT  02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT  02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE  GM2 activator protein.
OS  Homo sapiens (human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=92008637; PubMed=1915857;
RA  Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA  Sandhoff K.;
RT  "Characterization of full-length cDNAs and the gene coding for the
RT  human GM2 activator protein."
RL  FEBS Lett. 289:260-264(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93277527; PubMed=8503891;
RA  Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA  Sandhoff K.;
RT  "Over-expression of a functionally active human GM2-activator protein
RT  in escherichia coli."
RL  Biochem. J. 292:571-576(1993).
DR  EMBL; X62078; CAA43994.1; -.
SQ  SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match      97.0%; Score 987; DB 2; Length 189;
Best Local Similarity 98.9%; Pred. No. 7.1e-82;
Matches 187; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  5  MQAPLLIALGILLATPAQAHLKPSQSSFSWDCNCFEGKDPAVIRSLTLEPDPIVPGNV 64
DB  1  MQAPLLIALGILLATPAQAHLKPSQSSFSWDCNCFEGKDPAVIRSLTLEPDPIVPGNV 60

QY  65  TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGCTFEHFCVDLMDLPTGEP 124
DB  61  TLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGCTFEHFCVDLMDLPTGEP 120

QY  125  CPEPLRTYGLPCHCPFKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKRLGCI 184
DB  121  CPEPLRTYGLPCHCPFKEGTYSLPKSEFAVPDLELPSWLTGNYRIESVLSGSKRLGCI 180

QY  185  KIAASLKGI 193
DB  181  KIAASLKGI 189

RESULT 4
Q8HXX6  PRELIMINARY; PRT; 190 AA.
AC  Q8HXX6;
DT  01-MAR-2003 (TrEMBLrel. 23, Created)
DT  01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Ganglioside GM2 activator.
GN  Name=gm2a;
OS  Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC  Cercopitheciinae; Macaca.
OX  NCBI_TaxID=9541;
RN  [1]
RP  SEQUENCE FROM N.A.

```

```

RC TISSUE=Brain cerebellum cortex;
RA Kusuda J., Osada N., Hida M., Sugano S., Hashimoto K.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB083313; BAC20592.1; -.
DR HSP; P17900; IG13.
DR InterPro; IPR003172; E1_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 190 AA; 20494 MW; 9F9582BBB75715C3 CRC64;

Query Match 89.1%; Score 907.5; DB 2; Length 190;
Best Local Similarity 89.1%; Pred. No. 1.3e-74;
Matches 172; Conservative 8; Mismatches 10; Indels 3; Gaps 1;

QY 1 MQSILMQAPVLIALLGILLIATPAQAHLKPKPSQLSPSSWNCDFEGKDPVIRSLTLEPDPDIV 60
Db 1 MQSILMQAPVLIALLGILLIATPAQAHLKPKPSQLSPSSWNCDFEGKDPVIRSLTLEPDPDIV 57

QY 61 PGNVTVSVGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHCDFVLDMLIP 120
Db 58 PGNVTVSVGTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHCDFVLDMLIP 117

QY 121 TGFCPCPEPLRTYGLPCHCPKPKGTYSLPKSEFAVPDLELPSWLTGNYRIEVSLSGSKR 180
Db 118 TGFCPCPEPLRTYGLPCHCPKPKGTYSLPKSEFAVPDLELPSWLTGNYRIEVSLSGSKR 177

QY 181 LGCIKIAAASLKI 193
Db 178 LGCIKIAAASLKG 190

RESULT 5
Q6IN37 PRELIMINARY; PRT; 199 AA.
AC Q6IN37;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE GM2 ganglioside activator protein.
GN Names=gm2a;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; Pubmed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield V.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RN SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC072474; AAH72474.1; -.

```

```

DR InterPro; IPR003172; E1_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 199 AA; 21493 MW; C2B5203FDFABF507 CRC64;

Query Match 72.6%; Score 739.5; DB 2; Length 199;
Best Local Similarity 67.2%; Pred. No. 2.8e-59;
Matches 133; Conservative 24; Mismatches 32; Indels 9; Gaps 1;

QY 4 LMQAPLIIALLGILLIATPAQAHLKPKPSQLSPSSWNCDFEGKDPVIRSLTLE 54
Db 1 MRRVPLLVLGLLFLVGLLFAVPASRLSKPSQLGGFSWNCDFEGKDPVIRSLTLE 60

QY 55 PDPVIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHCDFV 114
Db 61 PDPVIVPGNVDIVSAGKTSIPLTSPOKVELTVEKAGFWVKIPCVQLGSCTYENVCDL 120

QY 115 LDMLIPTGPEPCPEPLRTYGLPCHCPKPKGTYSLPKSEFAVPDLELPSWLTGNYRIEVS 174
Db 121 IDQYIPPGTCPEPLHTYGLPCHCPKPKGTYSLPSSNFTVPDLELPSWLTGNYRIOSIL 180

QY 175 SSGKRLGCIKIAAASLKG 192
Db 181 SSGKRLACIKIAAASLKG 198

RESULT 6
Q8CJH4 PRELIMINARY; PRT; 199 AA.
AC Q8CJH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GM2 activator protein.
GN Names=rgm2ap;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RA Miwa N., Okada T., Nakamura S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051391; BAC24018.1; -.
DR HSP; P17900; IG13.
DR InterPro; IPR003172; E1_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 199 AA; 21521 MW; D585203FDFABF507 CRC64;

Query Match 72.6%; Score 739.5; DB 2; Length 199;
Best Local Similarity 67.2%; Pred. No. 2.8e-59;
Matches 133; Conservative 24; Mismatches 32; Indels 9; Gaps 1;

QY 4 LMQAPLIIALLGILLIATPAQAHLKPKPSQLSPSSWNCDFEGKDPVIRSLTLE 54
Db 1 MRRVPLLVLGLLFLVGLLFAVPASRLSKPSQLGGFSWNCDFEGKDPVIRSLTLE 60

QY 55 PDPVIVPGNVTLSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHCDFV 114
Db 61 PDPVIVPGNVDIVSAGKTSIPLTSPOKVELTVEKAGFWVKIPCVQLGSCTYENVCDL 120

QY 115 LDMLIPTGPEPCPEPLRTYGLPCHCPKPKGTYSLPKSEFAVPDLELPSWLTGNYRIEVS 174
Db 121 IDQYIPPGTCPEPLHTYGLPCHCPKPKGTYSLPSSNFTVPDLELPSWLTGNYRIOSIL 180

QY 175 SSGKRLGCIKIAAASLKG 192
Db 181 SSGKRLACIKIAAASLKG 198

RESULT 7
SAP3 MOUSE
ID SAP3 MOUSE STANDARD; PRT; 193 AA.
AC Q60648; Q61610; Q61819;

```


Db 1 KDPFAVIRSLTEPDPPIVPGNLTVSVEARTEVPLTSPQKVELTVEKEVAGFWAKVPCVEQ 60
QY 103 IGSCTFEHCDFVLDMLIPTGCPPEPLRTYGLPCHCFKGTYSLPKSEFAVPDLELPWS 162
Db 61 IGSCTYEDFCQIDTVTPPGCPPEPLHTYGLPCHCFKAGVYSLPESDFTLPQLEVPGW 120

QY 163 LTGTGNRYIESVSSGKRLGCIKIAASLKG 192
Db 121 LSSGHYRIKTVLSSGGERLGCVKISASLKG 150

RESULT 9

AA564350 PRELIMINARY; PRT; 151 AA.
ID AA564350;
AC AA564350;
DT 25-MAR-2004 (TrEMBLrel. 27, Created)
DT 25-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE GM2 activator protein (Fragment).
GN GM2A.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
RA Dodson A.N., Gentry A.S., Griffin B., Rolsma M.D., Baker H.J.;
RT "Mutation of the GM2 activator protein in a feline model of GM2
gangliosidosis".
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY553653; AA564350.1; -;
FT NON TER 1
SQ SEQUENCE 151 AA; 16310 MW; E23EF16CE5844A5F CRC64;

Query Match 59.1%; Score 602; DB 2; Length 151;
Best Local Similarity 68.7%; Pred. No. 6.9e-47;
Matches 103; Conservative 24; Mismatches 23; Indels 0; Gaps 0;

QY 43 KDPFAVIRSLTEPDPPIVPGNLTVSVEARTEVPLTSPQKVELTVEKEVAGFWAKVPCVEQ 102
Db 1 KDPFAVIRSLTEPDPPIVPGNLTVSVEARTEVPLTSPQKVELTVEKEVAGFWAKVPCVEQ 60
QY 103 IGSCTFEHCDFVLDMLIPTGCPPEPLRTYGLPCHCFKGTYSLPKSEFAVPDLELPWS 162
Db 61 IGSCTYEDFCQIDTVTPPGCPPEPLHTYGLPCHCFKAGVYSLPESDFTLPQLEVPGW 120

QY 163 LTGTGNRYIESVSSGKRLGCIKIAASLKG 192
Db 121 LSSGHYRIKTVLSSGGERLGCVKISASLKG 150

RESULT 10

Q14427 PRELIMINARY; PRT; 103 AA.
ID Q14427;
AC Q14427;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GM2-activator protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUP=Placenta;
RX MEDLINE=92207171; PubMed=1554364;
RA Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
RT "Evidence for two cDNAs encoding human GM2-activator protein.";
RL Biochem. J. 282:807-813(1992).
DR EMBL; X61094; CAA43407.1; -;

FT NON TER 1
SQ SEQUENCE 103 AA; 11137 MW; 434BFC334E976F91 CRC64;
Query Match 53.4%; Score 544; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 8.7e-42;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 KDPFAVIRSLTEPDPPIVPGNLTVSVEARTEVPLTSPQKVELTVEKEVAGFWAKVPCVEQ 102
Db 1 KDPFAVIRSLTEPDPPIVPGNLTVSVEARTEVPLTSPQKVELTVEKEVAGFWAKVPCVEQ 60

QY 103 IGSCTFEHCDFVLDMLIPTGCPPEPLRTYGLPCHCFK 142
Db 61 IGSCTFEHCDFVLDMLIPTGCPPEPLRTYGLPCHCFK 100

RESULT 11

Q6Q7X4 PRELIMINARY; PRT; 146 AA.
ID Q6Q7X4;
AC Q6Q7X4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE GM2 activator protein variant AB (Fragment).
GN Name=GM2A;
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,
RA Dodson A.N., Gentry A.S., Griffin B., Rolsma M.D., Baker H.J.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY553654; AA564351.1; -;
DR InterPro; IPR003172; EI_DerP2_DerF2.
DR SMART; SM00737; ML; 1;
FT NON TER 1
SQ SEQUENCE 146 AA; 15820 MW; 754E06C0CFCA74E3 CRC64;

Query Match 51.0%; Score 519; DB 2; Length 146;
Best Local Similarity 66.9%; Pred. No. 2.5e-39;
Matches 87; Conservative 21; Mismatches 22; Indels 0; Gaps 0;

QY 43 KDPFAVIRSLTEPDPPIVPGNLTVSVEARTEVPLTSPQKVELTVEKEVAGFWAKVPCVEQ 102
Db 1 KDPFAVIRSLTEPDPPIVPGNLTVSVEARTEVPLTSPQKVELTVEKEVAGFWAKVPCVEQ 60
QY 103 IGSCTFEHCDFVLDMLIPTGCPPEPLRTYGLPCHCFKGTYSLPKSEFAVPDLELPWS 162
Db 61 IGSCTYEDFCQIDTVTPPGCPPEPLHTYGLPCHCFKAGVYSLPESDFTLPQLEVPGW 120

QY 163 LTGTGNRYIES 172
Db 121 LSSGHYRIKT 130

RESULT 12

AA564351 PRELIMINARY; PRT; 146 AA.
ID AA564351;
AC AA564351;
DT 25-MAR-2004 (TrEMBLrel. 27, Created)
DT 25-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE GM2 activator protein variant AB (Fragment).
GN GM2A.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RA Martin D.R., Cox N.R., Morrison N.E., Kennamer D.M., Peck S.L.,

SEQUENCE FROM
LN
RN
PB

..Lrel. 27, Created)
..TEmBLrel. 27, Last sequence update)
..4 (TEmBLrel. 27, Last annotation update)
..C-1.
C Lymnaea stagnalis (Great pond snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
OC Lymnaeidae; Lymnaeidae; Lymnaea.
OX NCBI_TaxID=6523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14986137;
RA Harada Y., Hosoiri Y., Kuroda R.;
RT "Isolation and evaluation of dextral-specific and dextral-enriched
RT cDNA clones as candidates for the handedness-determining gene in a
RT freshwater gastropod, Lymnaea stagnalis.";
RL Dev. Genes Evol. 214:159-169(2004).
DR EMBL; AB159153; BADI601.1; -.
SQ SEQUENCE 217 AA; 23436 MW; DD626A73362E38AB CRC64;
Query Match 25.2%; Score 257; DB 2; Length 217;
Best Local Similarity 36.5%; Pred. No. 3e-15;
Matches 58; Conservative 22; Mismatches 69; Indels 10; Gaps 4;
QY 33 SFSWDCNCFEGKDPVIRSLTLEDDPIVVPNGVTLVVVGSTSVLSPLKVDLVLEKEVA 91
|||: |||: : : ||||| ||||| ||||| : ||| : |||
Db 67 SFSFKNCADPDNEILVPSNFNLEPDPIRAPGNITVSGNLEIKSKFGSPLVASVVVWKKVL 126
|||: |||: : : ||||| ||||| ||||| : ||| : |||
QY 92 GLWIKICTDYIGSTPEHFCDVLDMLIPTGCEPCPEPLRTYGLPCHCPKEGTYSLPKSE 151
|||: |||: : ||||| : |||: ||||| ||||| ||||| : ||| : |||
Db 127 GIWIKIPCYHGVGCTVSDACTLL-----TSPDCPTVLTGILPCCQCPFPAGTFNFBFFD 181
|||: |||: : ||||| : |||: ||||| ||||| ||||| : ||| : |||
QY 152 FAVPDLPLPSMLTTGNYRIESVLSSSGKRLGCIKIAASL 190
|||: |||: : ||||| : |||: ||||| ||||| ||||| : ||| : |||
Db 182 IVIPK-ALP---VSGEIFHLKTSVEGSLVTCVDLQFEL 216
|||: |||: : ||||| : |||: ||||| ||||| ||||| : ||| : |||

Search completed: November 17, 2004, 06:33:17
Job time : 30.6589 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 09:12:54 ; Search time 4.68912 Seconds
(without alignments)
1224.039 Million cell updates/sec

Title: US-10-030-937-68
Perfect score: 16
Sequence: 1 FSWDNCFEGKDPVIR 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A.Geneseq_23Sep04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	100.0	16	AAB31926	Aab31926 Amino aci
2	16	100.0	193	AAB31897	Aab31897 Amino aci
3	9	56.2	178	AAB31898	Aab31898 Amino aci
4	9	56.2	178	ABG31346	Abg31346 Non-glyco
5	9	56.2	189	AAB31900	Aab31900 Amino aci
6	9	56.2	193	AAB31901	Aab31901 Amino aci
7	9	56.2	193	AAB31902	Aab31902 Amino aci
8	9	56.2	193	AAB31904	Aab31904 Amino aci
9	9	56.2	193	AAB31896	Aab31896 Amino aci
10	9	56.2	193	AAB31928	Aab31928 Amino aci
11	9	56.2	193	AAB31903	Aab31903 Amino aci
12	9	56.2	193	ABG00720	Abg00720 Novel hum
13	9	56.2	193	ABG31345	Abg31345 Human GM2
14	9	56.2	193	ABP65212	Abp65212 Hypoxia-r
15	9	56.2	193	ADN95858	Adn95858 Human BEC
16	9	56.2	193	ADN03620	Adn03620 Antipsori
17	9	56.2	193	ADQ17712	Adq17712 Human sof
18	9	56.2	193	AAB31899	Aab31899 Amino aci
19	8	50.0	199	AAW10656	Aaw10656 Rac GM2 a
20	6	37.5	53	AAAB6944	Aaam6944 Human imm
21	6	37.5	85	ABP33559	Abp33559 Human ORF
22	6	37.5	91	ABG06658	Abg06658 Novel hum
23	6	37.5	91	ABP06013	Abp06013 Human ORF
24	6	37.5	93	ABB17405	Abb17405 Human ner
25	6	37.5	118	ABP64938	Abp64938 Human pro

26	6	37.5	126	4	AAU33628	Aau33628 Pseudomon
27	6	37.5	126	6	ABU15503	Abu15503 Protein e
28	6	37.5	127	7	ABO69666	Abo69666 Pseudomon
29	6	37.5	143	4	AAU95402	Aau95402 Human rep
30	6	37.5	143	4	ABB96093	Abb96093 Human tes
31	6	37.5	168	4	ABG06655	Abg06655 Novel hum
32	6	37.5	170	5	ABB81217	Abb81217 Human bra
33	6	37.5	170	6	AAE37969	Aae37969 Human kin
34	6	37.5	182	4	AAU48256	Aau48256 Propionib
35	6	37.5	182	6	ABM44775	Abm44775 Propionib
36	6	37.5	196	7	ADM26885	Adm26885 Hyperther
37	6	37.5	199	7	ADF74794	Adf74794 Human NOV
38	6	37.5	203	7	ADF74792	Adf74792 Human NOV
39	6	37.5	217	7	ADF74790	Adf74790 Human NOV
40	6	37.5	243	4	AAE13072	Aae13072 Homo sapi
41	6	37.5	243	7	ADD13128	Add13128 Human TPX
42	6	37.5	243	7	ADF74788	Adf74788 Human NOV
43	6	37.5	243	8	ADE94307	Ade94307 Human CYS
44	6	37.5	252	8	ADM16896	Adm16896 PRL-P6-El
45	6	37.5	257	4	ABG06656	Abg06656 Novel hum

ALIGNMENTS

RESULT 1
AAB31926
ID AAB31926 standard; peptide; 16 AA.
XX
XX AAB31926;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a peptide fragment of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
FN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR) BIOMERIEUX STELHYS.
XX
PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX
DR WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand.
XX
PS Claim 24; Page 108; 209pp; French.
XX
CC The present sequence represents a peptide fragment of a human protein,
XX which is used in the method of the invention. The specification describes
XX a method which uses at least one polypeptide or polynucleotide sequence
XX belonging to the perlecan, precursor of the retinol-binding plasma
XX protein, precursor of the ganglioside GM2 activator, calgranulin B or
XX saposin B protein families. The method is used for detecting, preventing
XX or treating a degenerative, neurological and/or auto-immune disease. The
XX polynucleotides and polypeptides are used for diagnosis, prognosis,
XX prevention and treatment of multiple sclerosis (in its various forms and
XX phases). They may also be useful in cases of e.g. Alzheimer's and

CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 16 AA;

Query Match 100.0%; Score 16; DB 4; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1.1e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSWDNCFEKDPFAVIR 16
 |||||
 DB 1 FSWDNCFEKDPFAVIR 16

RESULT 2
 AAB31897
 ID AAB31897 standard; protein; 193 AA.

XX AAB31897;

AC AAB31897;

XX 15-MAY-2001 (first entry)

DE Amino acid sequence of a mutant ganglioside GM2 activator protein.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX Homo sapiens.

XX WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INNR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX Claim 25; Page 159-160; 209pp; French.

XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 193 AA;

Query Match 100.0%; Score 16; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 9.2e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FSWDNCFEKDPFAVIR 16
 |||||
 DB 34 FSWDNCFEKDPFAVIR 49

RESULT 3

AAB31898

ID AAB31898 standard; protein; 178 AA.

XX AAB31898;

XX 15-MAY-2001 (first entry)

XX Amino acid sequence of a human protein.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX Homo sapiens.

XX WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INNR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX Claim 1; Page 160; 209pp; French.

XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 178 AA;

Query Match 56.2%; Score 9; DB 4; Length 178;
 Best Local Similarity 100.0%; Pred. No. 0.038;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDPFAVIR 16
 |||||
 DB 26 EGKDPFAVIR 34

Best Local Similarity 100.0%; Pred. No. 0.038;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 EGKDPVAVIR 16
 |||||
Db 26 EGKDPVAVIR 34

RESULT 5
AAB31900
ID AAB31900 standard; protein; 189 AA.
XX AC AAB31900;
XX DT 15-MAY-2001 (first entry)
XX DE Amino acid sequence of a human protein.
XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX OS Homo sapiens.
XX PN WO200105422-A2.
XX PD 25-JAN-2001.
XX PF 17-JUL-2000; 2000WO-FR002057.
XX PR 15-JUL-1999; 99FR-00009372.
XX PA (INMR) BIOMERIEUX STELHYS.
XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX DR
XX PT Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand.
XX PS Claim 1; Page 161-162; 209pp; French.
XX CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX SQ Sequence 189 AA;

Query Match 56.2%; Score 9; DB 4; Length 189;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 6
AAB31901
ID AAB31901 standard; protein; 193 AA.
XX
AC AAB31901;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 162-163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 56.2%; Score 9; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDPVAVIR 16
Db |||||
41 EGKDPVAVIR 49

RESULT 7
AAB31902
ID AAB31902 standard; protein; 193 AA.
XX
AC AAB31902;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 162-163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 56.2%; Score 9; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDPVAVIR 16
Db |||||
41 EGKDPVAVIR 49

RESULT 8
AAB31904
ID AAB31904 standard; protein; 193 AA.
XX
AC AAB31904;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX

```

OS Homo sapiens.
 XX WO200105422-A2.
 PN XX
 PD XX
 PF 25-JAN-2001.
 XX
 PP 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 XX
 XX Detecting, preventing and treating degenerative, neurological and
 XX autoimmune diseases, particularly multiple sclerosis, using specified
 XX polypeptides or related nucleic acid or ligand.
 PS Claim 1; Page 164-165; 209pp; French.
 XX
 CC The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;
 Query Match 56.2%; Score 9; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 0.041;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 ECKDPAVIR 16
 Db |||||
 41 ECKDPAVIR 49
 RESULT 9
 AAB31896
 ID AAB31896 standard; protein; 193 AA.
 AC AAB31896;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of the human ganglioside GM2 activator protein.
 XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PP 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 XX

PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 DR N-PSDB; AAF54698.
 XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 23; Page 158-159; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;
 Query Match 56.2%; Score 9; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 0.041;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 ECKDPAVIR 16
 Db |||||
 41 ECKDPAVIR 49
 RESULT 10
 AAB31928
 ID AAB31928 standard; protein; 193 AA.
 XX
 AC AAB31928;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of the human ganglioside GM2 activator protein.
 XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PP 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 XX

XX Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX PS Disclosure; Fig 1; 209pp; French.
XX CC The present sequence represents a human polypeptide, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX CC Sequence 193 AA;
SQ

Query Match 56.2%; Score 9; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDPVAVR 16
| | | | | | | |
Db 41 EGKDPVAVR 49

RESULT 11
AAB31903
ID AAB31903 standard; protein; 193 AA.
XX CC AAB31903;
XX DT 15-MAY-2001 (first entry)
XX DE Amino acid sequence of a human protein.
XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
XX KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
XX KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
XX KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
XX KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX OS Homo sapiens.
XX PN WO200105422-A2.
XX PD 25-JAN-2001.
XX PF 17-JUL-2000; 2000WO-FR002057.
XX PR 15-JUL-1999; 99FR-00009372.
XX PA (INNR) BIOMERIEUX SPELHYS.
XX PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX DR Detecting, preventing and treating degenerative, neurological and
XX PT autoimmune diseases, particularly multiple sclerosis, using specified
XX PT polypeptides or related nucleic acid or ligand.
XX PS Claim 1; Page 164; 209pp; French.
XX CC The present sequence represents a human protein, which is used in the

CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX CC Sequence 193 AA;
SQ

Query Match 56.2%; Score 9; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDPVAVR 16
| | | | | | | |
Db 41 EGKDPVAVR 49

RESULT 12
ABG00720
ID ABG00720 standard; protein; 193 AA.
XX CC ABG00720;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #711.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS64907.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 31079; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 193 AA;

Query Match 56.2%; Score 9; DB 4; Length 193;

Best Local Similarity 100.0%; Pred. No. 0.041;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EKGKDAVIR 16

Db 41 EKGKDAVIR 49

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

CC autoimmune disease (such as lupus), hypersensitivity infection,
 CC rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
 CC gout and tissue-specific conditions (e.g. glomerulonephritis and
 CC hepatitis). The composition is capable of inhibiting platelet activating
 CC factor, is non-toxic, is efficacious and presents less severe side
 CC effects, including redness, swelling, pain and polymorphonuclear
 CC leukocyte accumulation at the inflammatory site and other associated
 CC cellular responses. The present sequence represents human GM2 activator
 CC protein

XX Sequence 193 AA;

Query Match 56.2%; Score 9; DB 5; Length 193;

Best Local Similarity 100.0%; Pred. No. 0.041;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EKGKDAVIR 16

Db 41 EKGKDAVIR 49

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

RESULT 14

ABP65212

ID ABP65212 standard; protein; 193 AA.

XX AC ABP65212;

XX DT 12-NOV-2002 (first entry)

XX DE Hypoxia-regulated protein #86.

XX KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;

XX KW antiinflammatory; vulnary; gynecological; ophthalmological; vaccine;

XX KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;

XX KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;

XX KW presclapmsia; atherosclerosis; inflammatory condition; wound healing;

XX KW inflammation; erythropoiesis; hair loss; human.

XX OS Homo sapiens.

XX PN WO200246465-A2.

XX PD 13-JUN-2002.

XX PF 10-DEC-2001; 2001WO-GB005458.

XX PR 08-DEC-2000; 2000GB-00030076.

XX PR 08-FEB-2001; 2001GB-00003156.

XX PR 25-OCT-2001; 2001GB-00025666.

XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX PI White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;

XX PI Rayner WN;

XX DR WPI; 2002-627238/67.

XX PT Identifying a gene involved in disease for treating hypoxia-regulated

XX PT conditions, comprises comparing the transcriptome/proteome of two cell

XX PT types under different conditions and identifying a differentially

XX PT regulated gene.

XX PS Claim 35; Page 397; 538pp; English.

XX CC The present invention relates to methods for identifying genes and

XX CC proteins that are implicated in a specific disease or physiological

XX CC condition. The method comprises comparing the transcriptome/proteome of a

XX CC specialised cell type implicated in a disease or condition with that of a

XX CC second specialised cell type, under two experimental conditions, and

XX CC identifying a gene that is differentially regulated in the two

XX CC specialised cell types under experimental conditions. ABV77873-ABV78116

XX CC and ABP65061-ABP65257 were identified using the methods of the invention.

XX CC The coding sequences and proteins are useful for treating a disease in a

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 193 AA;

Query Match 56.2%; Score 9; DB 4; Length 193;

Best Local Similarity 100.0%; Pred. No. 0.041;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EKGKDAVIR 16

Db 41 EKGKDAVIR 49

|||||

|||||

|||||

|||||

|||||

|||||

CC The present invention relates to a composition comprising GM2
 CC (ganglioside) activator protein or a GM2 activator peptide derived from
 CC the GM2 activator protein in combination with a carrier. The composition
 CC comprises the protein or peptide in an amount of 1-100 mg. The GM2
 CC activator protein is capable of inhibiting platelet activating factor
 CC (PAF). The composition of the invention is useful for treating
 CC inflammatory disorders e.g. inflammatory bowel disease, asthma,

XX Claim 3; Fig 1; 11pp; English.

XX PS The present invention relates to a composition comprising GM2

XX CC (ganglioside) activator protein or a GM2 activator peptide derived from

XX CC the GM2 activator protein in combination with a carrier. The composition

XX CC comprises the protein or peptide in an amount of 1-100 mg. The GM2

XX CC activator protein is capable of inhibiting platelet activating factor

XX CC (PAF). The composition of the invention is useful for treating

XX CC inflammatory disorders e.g. inflammatory bowel disease, asthma,

XX CC

XX CC

XX CC

XX CC

XX CC

CC patient, for manufacture of a medicament for treating hypoxia-regulated
 CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
 CC biological response to hypoxia conditions, or hypoxic-associated
 CC pathology in a patient. The coding sequences and proteins are also useful
 CC for monitoring the therapeutic treatment of a disease or physiological
 CC condition, such as cancer, ischaemic conditions, reperfusion injury,
 CC retinopathy, neonatal stress, preclampsia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss
 XX
 CC Sequence 193 AA;

Query Match 56.2%; Score 9; DB 5; Length 193;
 Best Local Similarity 100.0%; Pred. No. 0.041;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 EGKDPVAVIR 16
 |||||
 Db 41 EGKDPVAVIR 49

RESULT 15

ADN95858
 ID ADN95858 standard; protein; 193 AA.

AC ADN95858;

DT 01-JUL-2004 (first entry)

DE Human BEC/LEC-related protein sequence SeqID782.

XX growth; differentiation; blood endothelial cell; BEC;
 KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
 KW lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;
 KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
 KW inflammatory disease; cancer metastasis; lymphatic system; human.
 XX

OS Homo sapiens.

XX WO2003080640-A1.

XX 02-OCT-2003.

XX 07-MAR-2003; 2003WO-US006900.

XX 07-MAR-2002; 2002US-0363019P.

PR (LUDW-) LUDWIG INST CANCER RES.
 PA (LICN) LICENTIA LTD.

XX Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;

XX WPI; 2003-876899/81.

XX N-PSDB; ADN95859.

PS Example 1; SEQ ID NO 782; 176pp; English.

XX This invention relates to a method of differentially modulating the
 CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
 CC endothelial cells (LEC) comprises contacting endothelial cells with a
 CC composition comprising an agent that differentially modulates blood or
 CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
 CC identifying a human subject with lymphoedema and with a mutation in at
 CC least one allele of a gene encoding a LEC protein, where the mutation
 CC correlates with lymphoedema in human subjects, and with the proviso that
 CC the LEC protein is not VEGFR-3; and administering to the subject a
 CC composition comprising a lymphatic growth agent selected from VEGF-C or
 CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
 CC the development of compounds with an antiangiogenic, cytostatic,
 CC vasotropic or antiinflammatory activity or for gene therapy. The method
 CC is useful in modulating the growth or differentiation of blood
 CC endothelial cells or lymphatic endothelial cells, in treating hereditary
 CC lymphoedema, in screening for an endothelial cell disorder or
 CC predisposition to the disorder or in monitoring the efficacy or toxicity

CC of a drug on endothelial cells. The agent is useful in manufacturing a
 CC medicament for the differential modulation of blood vessel endothelial
 CC cell or lymphatic vessel endothelial cell growth or differentiation. The
 CC lymphatic growth agent may also be used in manufacturing a medicament for
 CC the treatment of hereditary lymphoedema resulting from a mutation in a
 CC LEC gene or of other diseases involving the lymphatic vessels, such as
 CC various inflammatory diseases and cancer metastasis via the lymphatic
 CC system. The present sequence is that of a human LEC/BEC differentially
 CC expressed protein which is related to the method of the invention. Note:
 CC This sequence does not appear in the specification but was obtained by
 CC the indexer using the source data given in table 14 of the specification.

XX Sequence 193 AA;

Query Match 56.2%; Score 9; DB 7; Length 193;

Best Local Similarity 100.0%; Pred. No. 0.041;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 EGKDPVAVIR 16
 |||||

Db 41 EGKDPVAVIR 49

Search completed: November 17, 2004, 10:32:59

Job time : 7.78003 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 10:53:27 ; Search time 1.14595 Seconds
(without alignments)
925.943 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 16

Sequence: 1 FSWNCFEGKDAVIR 16

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep:*

2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep:*

3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep:*

4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep:*

5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep:*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	56.2	178	4	US-09-183-841-2
2	9	56.2	193	4	US-09-183-841-1
3	6	37.5	127	4	US-09-252-991A-18412
4	6	37.5	185	4	US-09-248-796A-16879
5	6	37.5	243	4	US-09-541-759-7
6	6	37.5	297	1	US-08-534-910B-6
7	6	37.5	297	1	US-08-534-910B-7
8	6	37.5	297	1	US-08-534-910B-8
9	6	37.5	297	1	US-08-534-910B-9
10	6	37.5	297	1	US-08-534-910B-10
11	6	37.5	297	3	US-08-886-466-2
12	6	37.5	297	3	US-09-475-304-2
13	6	37.5	297	3	US-09-101-126-3
14	6	37.5	297	3	US-09-367-528A-1
15	6	37.5	297	3	US-09-367-528A-3
16	6	37.5	297	3	US-09-367-528A-5
17	6	37.5	1765	4	US-09-270-767-45587
18	5	31.2	25	1	US-07-706-699-1
19	5	31.2	25	1	US-07-998-931-1
20	5	31.2	25	4	US-09-042-460-52
21	5	31.2	30	3	US-08-851-843A-155
22	5	31.2	30	3	US-08-974-549A-275
23	5	31.2	30	3	US-08-854-050-155
24	5	31.2	30	3	US-09-430-323-155
25	5	31.2	30	4	US-09-402-181B-275
26	5	31.2	30	4	US-09-721-456-275
27	5	31.2	35	2	US-08-737-716-12

28 5 31.2 54 3 US-08-974-549A-25 Sequence 25, Appl
29 5 31.2 54 4 US-08-912-951-25 Sequence 25, Appl
30 5 31.2 54 4 US-09-402-181B-25 Sequence 25, Appl
31 5 31.2 54 4 US-09-721-456-25 Sequence 25, Appl
32 5 31.2 64 4 US-09-205-258-774 Sequence 774, App
33 5 31.2 78 4 US-09-489-039A-11104 Sequence 11104, A
34 5 31.2 82 4 US-09-198-452A-1142 Sequence 1142, Ap
35 5 31.2 82 4 US-09-270-767-58568 Sequence 58568, A
36 5 31.2 82 4 US-09-513-999C-5870 Sequence 5870, Ap
37 5 31.2 107 4 US-09-489-039A-11962 Sequence 11962, A
38 5 31.2 127 4 US-09-270-767-47503 Sequence 47503, A
39 5 31.2 145 4 US-09-270-767-43226 Sequence 43226, A
40 5 31.2 149 4 US-09-270-767-40911 Sequence 40911, A
41 5 31.2 170 4 US-09-732-210-558 Sequence 558, App
42 5 31.2 180 4 US-09-149-476-401 Sequence 401, App
43 5 31.2 181 3 US-09-129-030-8 Sequence 8, Appli
44 5 31.2 181 4 US-09-443-067-10 Sequence 10, Appl
45 5 31.2 181 4 US-09-443-067-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-183-841-2
; Sequence 2, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: His tag at residues 1 to 17
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: amino acid
; OTHER INFORMATION: sequence of GM2 protein using His6 tag
US-09-183-841-2

Query Match 56.2%; Score 9; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 0.0089;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDAVIR 16
DB 26 EGKDAVIR 34

RESULT 2

US-09-183-841-1
; Sequence 1, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL

```
; LOCATION: (33)...(55)
; FEATURE:
; OTHER INFORMATION: residues 56-63 are included in a further precursor
; OTHER INFORMATION: form of the protein
US-09-183-841-1

Query Match          56.2%; Score 9; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.0095;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 EGKDPVAVR 16
      |||||
Db      41 EGKDPVAVR 49

RESULT 3
US-09-252-991A-18412
; Sequence 18412, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18412
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18412

Query Match          37.5%; Score 6; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 EGKDKPA 13
      |||||
Db      4 EGKDKPA 9

RESULT 4
US-09-248-796A-16879
; Sequence 16879, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16879
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16879

Query Match          37.5%; Score 6; DB 4; Length 185;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      10 KDPVAVI 15
```

```
Db      171 KDPVAVI 176
      |||||

RESULT 5
US-09-541-759-7
; Sequence 7, Application US/09541759
; Patent No. 6723322
; GENERAL INFORMATION:
; APPLICANT: Lustigman, Sara
; APPLICANT: Pearlman, Eric
; APPLICANT: Unnasch, Thomas
; TITLE OF INVENTION: ANGIOGENIC ONCHOCERCA VOLVULUS PROTEINS AND USES THEREOF
; FILE REFERENCE: 63475/252
; CURRENT APPLICATION NUMBER: US/09/541.759
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-541-759-7

Query Match          37.5%; Score 6; DB 4; Length 243;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 EGKDKPA 13
      |||||
Db      20 EGKDKPA 25

RESULT 6
US-08-534-910B-6
; Sequence 6, Application US/08534910B
; Patent No. 5766911
; GENERAL INFORMATION:
; APPLICANT: KOIKE, Ayumi
; APPLICANT: OBATA, Shusei
; APPLICANT: NISHINO, Tokuzo
; APPLICANT: OHNUMA, Shinichi
; APPLICANT: NAKAZAWA, Takeshi
; APPLICANT: OGURA, Kyoza
; APPLICANT: KOYAMA, Tanetoshi
; TITLE OF INVENTION: Mutated Farnesylidiphosphate Synthase Capable
; TITLE OF INVENTION: Of Synthesizing Geranylgeranyldiphosphate And Gene Coding Ther
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.
; ZIP: 20036-5405
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.25" Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: IBM/Word Perfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/534,910B
; FILING DATE: 28-SEPT-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-25253
; FILING DATE: 14-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 77670/398
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)429-1776
```


TELEFAX: (202)429-0796
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacillus stearothermophilus
US-08-534-910B-6

Query Match 37.5%; Score 6; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GKDPV 14
Db 62 GKDPV 67

RESULT 7
US-08-534-910B-7
; Sequence 7, Application US/08534910B
; Patent No. 5766911
; GENERAL INFORMATION:
; APPLICANT: KOIKE, Ayumi
; APPLICANT: OBATA, Shusei
; APPLICANT: NISHINO, Tokuzo
; APPLICANT: OHNUMA, Shinichi
; APPLICANT: NAKAZAWA, Takeshi
; APPLICANT: OGURA, Kyoza
; APPLICANT: KOYAMA, Tanetoshi
; TITLE OF INVENTION: Mutated Farnesylidiphosphate Synthase Capable
; TITLE OF INVENTION: Of Synthesizing Geranylgeranylidiphosphate And Gene Coding There
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.
; ZIP: 20036-5405
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.25" Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: IBM/Word Perfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/534,910B
; FILING DATE: 28-SEPT-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-25253
; FILING DATE: 14-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 77670/398
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)429-1776
; TELEFAX: (202)429-0796
; INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacillus stearothermophilus
US-08-534-910B-7

Query Match 37.5%; Score 6; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GKDPV 14
Db 62 GKDPV 67

RESULT 8
US-08-534-910B-8
; Sequence 8, Application US/08534910B
; Patent No. 5766911
; GENERAL INFORMATION:
; APPLICANT: KOIKE, Ayumi
; APPLICANT: OBATA, Shusei
; APPLICANT: NISHINO, Tokuzo
; APPLICANT: OHNUMA, Shinichi
; APPLICANT: NAKAZAWA, Takeshi
; APPLICANT: OGURA, Kyoza
; APPLICANT: KOYAMA, Tanetoshi
; TITLE OF INVENTION: Mutated Farnesylidiphosphate Synthase Capable
; TITLE OF INVENTION: Of Synthesizing Geranylgeranylidiphosphate And Gene Coding There
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.
; ZIP: 20036-5405
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.25" Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: IBM/Word Perfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/534,910B
; FILING DATE: 28-SEPT-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-25253
; FILING DATE: 14-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 77670/398
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)429-1776
; TELEFAX: (202)429-0796
; INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacillus stearothermophilus
US-08-534-910B-8

Query Match 37.5%; Score 6; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GKDPV 14
Db 62 GKDPV 67

RESULT 9
US-08-534-910B-9
; Sequence 9, Application US/08534910B
; Patent No. 5766911
; GENERAL INFORMATION:
; APPLICANT: KOIKE, Ayumi

```
;; APPLICANT: OBATA, Shusei
;; APPLICANT: NISHINO, Tokuzo
;; APPLICANT: OHNUMA, Shinichi
;; APPLICANT: NAKAZAWA, Takeshi
;; APPLICANT: OGURA, Kyoze
;; APPLICANT: KOYAMA, Tanetoshi
;; TITLE OF INVENTION: Mutated Farnesylidiphosphate Synthase Capable
;; TITLE OF INVENTION: Of Synthesizing Geranylgeranyldiphosphate And Gene Coding Ther
;; NUMBER OF SEQUENCES: 10
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Kenyon & Kenyon
;; STREET: 1025 Connecticut Avenue, N.W., Suite 600
;; CITY: Washington
;; STATE: DC
;; COUNTRY: U.S.
;; ZIP: 20036-5405
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 3.25" Floppy Disk
;; COMPUTER: IBM PC Compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
;; SOFTWARE: IBM/Word Perfect 6.1 Windows
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/534,910B
;; FILING DATE: 28-SEPT-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 7-25253
;; FILING DATE: 14-FEB-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Toffenetti, Judith L.
;; REGISTRATION NUMBER: 39,048
;; REFERENCE/DOCKET NUMBER: 77670/398
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)429-1776
;; TELEFAX: (202)429-0796
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 297 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; ORIGINAL SOURCE:
;; ORGANISM: Bacillus stearothermophilus
US-08-534-910B-9
Query Match 37.5%; Score 6; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9 GKDPV 14
Db 62 GKDPV 67
RESULT 10
US-08-534-910B-10
; Sequence 10, Application US/08534910B
; Patent No. 5766911
; GENERAL INFORMATION:
; APPLICANT: KOIKE, Ayumi
; APPLICANT: OBATA, Shusei
; APPLICANT: NISHINO, Tokuzo
; APPLICANT: OHNUMA, Shinichi
; APPLICANT: NAKAZAWA, Takeshi
; APPLICANT: OGURA, Kyoze
; APPLICANT: KOYAMA, Tanetoshi
; TITLE OF INVENTION: Mutated Farnesylidiphosphate Synthase Capable
; TITLE OF INVENTION: Of Synthesizing Geranylgeranyldiphosphate And Gene Coding Ther
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
```

```
;; STATE: DC
;; COUNTRY: U.S.
;; ZIP: 20036-5405
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 3.25" Floppy Disk
;; COMPUTER: IBM PC Compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
;; SOFTWARE: IBM/Word Perfect 6.1 Windows
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/534,910B
;; FILING DATE: 28-SEPT-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 7-25253
;; FILING DATE: 14-FEB-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Toffenetti, Judith L.
;; REGISTRATION NUMBER: 39,048
;; REFERENCE/DOCKET NUMBER: 77670/398
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)429-1776
;; TELEFAX: (202)429-0796
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 297 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; ORIGINAL SOURCE:
;; ORGANISM: Bacillus stearothermophilus
US-08-534-910B-10
Query Match 37.5%; Score 6; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9 GKDPV 14
Db 62 GKDPV 67
RESULT 11
US-08-886-466-2
; Sequence 2, Application US/08886466C
; Patent No. 6040165
; GENERAL INFORMATION:
; APPLICANT: Narita, Keishi
; APPLICANT: Ishida, Chika
; APPLICANT: Takeuchi, Yoshie
; APPLICANT: Ohto, Chikara
; APPLICANT: Ohnuma, Shinichi
; APPLICANT: Nishino, Tokuzo
; TITLE OF INVENTION: MUTANT PRENYL DIPHOSPHATE SYNTHASE
; FILE REFERENCE: 77670/494
; CURRENT APPLICATION NUMBER: US/08/886,466C
; CURRENT FILING DATE: 1997-07-10
; EARLIER APPLICATION NUMBER: JP 8-191635
; EARLIER FILING DATE: 1996-07-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-08-886-466-2
Query Match 37.5%; Score 6; DB 3; Length 297;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9 GKDPV 14
Db 62 GKDPV 67
```

RESULT 14
US-09-367-528A-1
; Sequence 1, Application US/09367528A
; Patent No. 6395525
; GENERAL INFORMATION:
; APPLICANT: TOYOTA JIDOSHA KABUSHIKI KAISHA
; TITLE OF INVENTION: Geranyl Diphosphate Synthetase Gene
; FILE REFERENCE: PH-586
; CURRENT APPLICATION NUMBER: US/09/367,528A
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: JP97/346686
; PRIOR FILING DATE: 1997-12-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: 82
; OTHER INFORMATION: Xaa represents Val, Leu, Ile, Thr, Asp, Glu, Asn, Gln, Lys,
; OTHER INFORMATION: Arg, Cys, Met, Phe, Tyr, Trp, His or Pro.
US-09-367-528A-1
Query Match 37.5%; Score 6; DB 3; Length 297;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 GKDPV 14
Db 62 GKDPV 67
RESULT 15
US-09-367-528A-3
; Sequence 3, Application US/09367528A
; Patent No. 6395525
; GENERAL INFORMATION:
; APPLICANT: TOYOTA JIDOSHA KABUSHIKI KAISHA
; TITLE OF INVENTION: Geranyl Diphosphate Synthetase Gene
; FILE REFERENCE: PH-586
; CURRENT APPLICATION NUMBER: US/09/367,528A
; CURRENT FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: JP97/346686
; PRIOR FILING DATE: 1997-12-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-367-528A-3
Query Match 37.5%; Score 6; DB 3; Length 297;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 GKDPV 14
Db 62 GKDPV 67

Search completed: November 17, 2004, 12:29:32
Job time : 2.23686 secs

RESULT 12
US-09-475-304-2
; Sequence 2, Application US/09475304
; Patent No. 6225096
; GENERAL INFORMATION:
; APPLICANT: Narita, Keishi
; APPLICANT: Ishida, Chika
; APPLICANT: Takeuchi, Yoshie
; APPLICANT: Ohto, Chikara
; APPLICANT: Ohnuma, Shinichi
; APPLICANT: Nishino, Tokuzo
; TITLE OF INVENTION: MUTANT PRENYL DIPHOSPHATE SYNTHASE
; FILE REFERENCE: 77670/494
; CURRENT APPLICATION NUMBER: US/09/475,304
; CURRENT FILING DATE: 1999-12-30
; EARLIER APPLICATION NUMBER: JP 8-191635
; EARLIER FILING DATE: 1996-07-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-475-304-2
Query Match 37.5%; Score 6; DB 3; Length 297;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 GKDPV 14
Db 62 GKDPV 67
RESULT 13
US-09-101-126-3
; Sequence 3, Application US/09101126
; Patent No. 6316216
; GENERAL INFORMATION:
; APPLICANT: OHTO, CHIKARA
; APPLICANT: NAKANE, HIROYUKI
; APPLICANT: NISHINO, TOKUZO
; APPLICANT: OHNUMA, SHINICHI
; APPLICANT: HIROOKA, KAZUTAKE
; TITLE OF INVENTION: MUTATED PRENYL DIPHOSPHATE SYNTHASES
; FILE REFERENCE: 77670/566
; CURRENT APPLICATION NUMBER: US/09/101,126
; CURRENT FILING DATE: 1999-04-27
; EARLIER APPLICATION NUMBER: PCT/JP97/03921
; EARLIER FILING DATE: 1997-10-29
; EARLIER APPLICATION NUMBER: JP 8-307506
; EARLIER FILING DATE: 1996-11-05
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
; FEATURE:
; OTHER INFORMATION: 86-92 is an Asp-rich domain
US-09-101-126-3
Query Match 37.5%; Score 6; DB 3; Length 297;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 GKDPV 14
Db 62 GKDPV 67

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 12:24:02 ; Search time 3.8405 Seconds
(without alignments)
1475.341 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 16

Sequence: 1 FSWNCFEGKDPVIR 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1570615 seqs, 354127592 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	56.2	193	14	US-10-170-385-389
2	6	37.5	85	11	US-09-864-408A-5064
3	6	37.5	105	15	US-10-424-599-274901
4	6	37.5	126	9	US-09-815-242-5124
5	6	37.5	126	15	US-10-282-122A-43427
6	6	37.5	143	10	US-09-764-891-4060
7	6	37.5	160	15	US-10-424-599-260239
8	6	37.5	199	15	US-10-383-201-8
9	6	37.5	203	15	US-10-383-201-6
10	6	37.5	217	15	US-10-383-201-4
11	6	37.5	243	15	US-10-383-201-2
12	6	37.5	297	9	US-09-367-528A-1
13	6	37.5	297	9	US-09-367-528A-3

Sequence 5, Appli
Sequence 76, Appl
Sequence 52133, A
Sequence 588, App
Sequence 60813, A
Sequence 200217,
Sequence 20145, A
Sequence 46, Appl
Sequence 49297, A
Sequence 203808,
Sequence 10, Appl
Sequence 3, Appli
Sequence 6665, Ap
Sequence 137331,
Sequence 185490,
Sequence 185490,
Sequence 137333,
Sequence 126411,
Sequence 137323,
Sequence 126407,
Sequence 126375,
Sequence 126488,
Sequence 20151, A
Sequence 631, App
Sequence 632, App
Sequence 633, App
Sequence 587, App
Sequence 588, App
Sequence 589, App
Sequence 590, App
Sequence 43155, A

ALIGNMENTS

RESULT 1
US-10-170-385-389
Sequence 389, Application US/10170385
Publication No. US20030203372A1
GENERAL INFORMATION:
APPLICANT: Ward, Neil Raymond
APPLICANT: Mundy, Neil Raymond
APPLICANT: Kan, On
APPLICANT: Harris, Robert Alan
APPLICANT: White, Jonathan
APPLICANT: Binley, Katie Mary
APPLICANT: Rayner, William Nigel
APPLICANT: Naylor, Stuart
APPLICANT: Kingsman, Susan Mary
APPLICANT: Krige, David
TITLE OF INVENTION: ANALYSIS METHOD
FILE REFERENCE: 532682000100
CURRENT APPLICATION NUMBER: US/10/170,385
CURRENT FILING DATE: 2002-06-12
PRIOR APPLICATION NUMBER: PCT/GB02/01662
PRIOR FILING DATE: 2002-04-08
PRIOR APPLICATION NUMBER: PCT/GB01/05458
PRIOR FILING DATE: 2001-12-10
NUMBER OF SEQ ID NOS: 549
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 389
LENGTH: 193
TYPE: PRT
ORGANISM: Homo Sapiens
US-10-170-385-389

Query Match 56.2%; Score 9; DB 14; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGDPAVIR 16

Db 41 EGKDPVIR 49
|||||

RESULT 2
US-09-864-408A-5064
; Sequence 5064, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encoded by Genes
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5064
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(1)
; OTHER INFORMATION: Wherein Xaa may be any naturally occurring amino acid
US-09-864-408A-5064

Query Match 37.5%; Score 6; DB 11; Length 85;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 DPAVIR 16
|||||
Db 73 DPAVIR 78

RESULT 3
US-10-424-599-274901
; Sequence 274901, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with Soy
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 274901
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(105)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_90256C.1.pep
US-10-424-599-274901

Query Match 37.5%; Score 6; DB 15; Length 105;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 EGKDDPA 13
|||||
Db 42 EGKDDPA 47

RESULT 4
US-09-815-242-5124
; Sequence 5124, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5124
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5124

Query Match 37.5%; Score 6; DB 9; Length 126;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 EGKDDPA 13
|||||
Db 3 EGKDDPA 8

RESULT 5
US-10-282-122A-43427
; Sequence 43427, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 43427
;; LENGTH: 126
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-43427

Query Match 37.5%; Score 6; DB 15; Length 126;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGDPA 13
|||
Db 3 EGDPA 8

RESULT 6
US-09-764-891-4060
;; Sequence 4060, Application US/09764891
;; Publication No. US2003007808A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC006
;; CURRENT APPLICATION NUMBER: US/09/764,891
;; PRIOR FILING DATE: 2001-01-17
;; Prior application data removed - consult PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 10231
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 4060
;; LENGTH: 143
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (65)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;; NAME/KEY: SITE
;; LOCATION: (126)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;; NAME/KEY: SITE
;; LOCATION: (131)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4060

Query Match 37.5%; Score 6; DB 10; Length 143;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGDPA 13
|||
Db 76 EGDPA 81

RESULT 7
US-10-424-599-260239
;; Sequence 260239, Application US/10424599
;; Publication No. US20040031072A1
;; GENERAL INFORMATION:
;; APPLICANT: La Rosa Thomas J
;; APPLICANT: Kovalic David K
;; APPLICANT: Zhou Yihua
;; APPLICANT: Cao Yongwei
;; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;; FILE REFERENCE: 38-21(53223)B
;; CURRENT APPLICATION NUMBER: US/10/424,599
;; CURRENT FILING DATE: 2003-04-28
;; NUMBER OF SEQ ID NOS: 285684
;; SEQ ID NO 260239
;; LENGTH: 160
;; TYPE: PRT
;; ORGANISM: Glycine max
;; FEATURE:
;; OTHER INFORMATION: Clone ID: PAT_MRT3847_77019C.1.pep
US-10-424-599-260239

Query Match 37.5%; Score 6; DB 15; Length 160;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DPAVIR 16
|||
Db 99 DPAVIR 104

RESULT 8
US-10-383-201-8
;; Sequence 8, Application US/10383201
;; Publication No. US20040029226A1
;; GENERAL INFORMATION:
;; APPLICANT: Alsobrook II, John et al.
;; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
;; FILE REFERENCE: 21402-568A
;; CURRENT APPLICATION NUMBER: US/10/383,201
;; CURRENT FILING DATE: 2003-03-06
;; PRIOR APPLICATION NUMBER: 10/029020
;; PRIOR FILING DATE: 2001-12-19
;; PRIOR APPLICATION NUMBER: 60/365,984
;; PRIOR FILING DATE: 2002-03-20
;; PRIOR APPLICATION NUMBER: 60/372,022
;; PRIOR FILING DATE: 2002-04-12
;; PRIOR APPLICATION NUMBER: 60/389,143
;; PRIOR FILING DATE: 2002-06-14
;; PRIOR APPLICATION NUMBER: 60/391,779
;; PRIOR FILING DATE: 2002-06-26
;; PRIOR APPLICATION NUMBER: 60/410,755
;; PRIOR FILING DATE: 2002-09-13
;; PRIOR APPLICATION NUMBER: 60/412,957
;; PRIOR FILING DATE: 2002-09-23
;; PRIOR APPLICATION NUMBER: 10/051,874
;; PRIOR FILING DATE: 2002-01-16
;; PRIOR APPLICATION NUMBER: 60/366,928
;; PRIOR FILING DATE: 2002-03-22
;; PRIOR APPLICATION NUMBER: 10/055,877
;; PRIOR FILING DATE: 2002-01-22
;; NUMBER OF SEQ ID NOS: 155
;; SOFTWARE: CuraSeqList version 0.1
;; SEQ ID NO 8
;; LENGTH: 199
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-383-201-8

Query Match 37.5%; Score 6; DB 15; Length 199;
Best Local Similarity 100.0%; Pred. No. 90;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKQPA 13
|||||
Db 2 EGKQPA 7

RESULT 9

US-10-383-201-6
; Sequence 6, Application US/10383201
; Publication No. US20040029226A1

; GENERAL INFORMATION:

; APPLICANT: Alsbrook II, John et al.

; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

; FILE REFERENCE: 21402-568A

; CURRENT APPLICATION NUMBER: US/10/383,201

; CURRENT FILING DATE: 2003-03-06

; PRIOR APPLICATION NUMBER: 10/029020

; PRIOR FILING DATE: 2001-12-19

; PRIOR APPLICATION NUMBER: 60/365,984

; PRIOR FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: 60/372,022

; PRIOR FILING DATE: 2002-04-12

; PRIOR APPLICATION NUMBER: 60/389,143

; PRIOR FILING DATE: 2002-06-14

; PRIOR APPLICATION NUMBER: 60/391,779

; PRIOR FILING DATE: 2002-06-26

; PRIOR APPLICATION NUMBER: 60/410,755

; PRIOR FILING DATE: 2002-09-13

; PRIOR APPLICATION NUMBER: 60/412,957

; PRIOR FILING DATE: 2002-09-23

; PRIOR APPLICATION NUMBER: 10/051,874

; PRIOR FILING DATE: 2002-01-16

; PRIOR APPLICATION NUMBER: 60/366,928

; PRIOR FILING DATE: 2002-03-22

; PRIOR APPLICATION NUMBER: 10/055,877

; PRIOR FILING DATE: 2002-01-22

; NUMBER OF SEQ ID NOS: 155

; SOFTWARE: CuraSeqList version 0.1

; SEQ ID NO 6

; LENGTH: 203

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-383-201-6

Query Match 37.5%; Score 6; DB 15; Length 203;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKQPA 13
|||||
Db 4 EGKQPA 9

RESULT 10

US-10-383-201-4
; Sequence 4, Application US/10383201
; Publication No. US20040029226A1

; GENERAL INFORMATION:

; APPLICANT: Alsbrook II, John et al.

; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

; FILE REFERENCE: 21402-568A

; CURRENT APPLICATION NUMBER: US/10/383,201

; CURRENT FILING DATE: 2003-03-06

; PRIOR APPLICATION NUMBER: 10/029020

; PRIOR FILING DATE: 2001-12-19

; PRIOR APPLICATION NUMBER: 60/365,984

; PRIOR FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: 60/372,022

; PRIOR FILING DATE: 2002-04-12

; PRIOR APPLICATION NUMBER: 60/389,143

; PRIOR FILING DATE: 2002-06-14

; PRIOR APPLICATION NUMBER: 60/391,779

; PRIOR FILING DATE: 2002-06-26
; PRIOR APPLICATION NUMBER: 60/410,755
; PRIOR FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 60/412,957
; PRIOR FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: 10/051,874
; PRIOR FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 60/366,928
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 10/055,877
; PRIOR FILING DATE: 2002-01-22
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 4
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-383-201-4

Query Match 37.5%; Score 6; DB 15; Length 217;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKQPA 13
|||||
Db 20 EGKQPA 25

RESULT 11

US-10-383-201-2

; Sequence 2, Application US/10383201

; Publication No. US20040029226A1

; GENERAL INFORMATION:

; APPLICANT: Alsbrook II, John et al.

; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

; FILE REFERENCE: 21402-568A

; CURRENT APPLICATION NUMBER: US/10/383,201

; CURRENT FILING DATE: 2003-03-06

; PRIOR APPLICATION NUMBER: 10/029020

; PRIOR FILING DATE: 2001-12-19

; PRIOR APPLICATION NUMBER: 60/365,984

; PRIOR FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: 60/372,022

; PRIOR FILING DATE: 2002-04-12

; PRIOR APPLICATION NUMBER: 60/389,143

; PRIOR FILING DATE: 2002-06-14

; PRIOR APPLICATION NUMBER: 60/391,779

; PRIOR FILING DATE: 2002-06-26

; PRIOR APPLICATION NUMBER: 60/410,755

; PRIOR FILING DATE: 2002-09-13

; PRIOR APPLICATION NUMBER: 60/412,957

; PRIOR FILING DATE: 2002-09-23

; PRIOR APPLICATION NUMBER: 10/051,874

; PRIOR FILING DATE: 2002-01-16

; PRIOR APPLICATION NUMBER: 60/366,928

; PRIOR FILING DATE: 2002-03-22

; PRIOR APPLICATION NUMBER: 10/055,877

; PRIOR FILING DATE: 2002-01-22

; NUMBER OF SEQ ID NOS: 155

; SOFTWARE: CuraSeqList version 0.1

; SEQ ID NO 2

; LENGTH: 243

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-383-201-2

Query Match

Best Local Similarity 37.5%; Score 6; DB 15; Length 243;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKQPA 13
|||||
Db 20 EGKQPA 25


```
RESULT 12
US-09-367-528A-1
; Sequence 1, Application US/09367528A
; Publication No. US20010051359A1
; GENERAL INFORMATION:
; APPLICANT: TOYOTA JIDOSHA KABUSHIKI KAISHA
; TITLE OF INVENTION: Geranyl Diphosphate Synthetase Gene
; FILE REFERENCE: PH-586
; CURRENT APPLICATION NUMBER: US/09/367,528A
; CURRENT FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: JP97/346686
; PRIOR FILING DATE: 1997-12-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
; NAME/KEY: PEPTIDE
; LOCATION: 82
; OTHER INFORMATION: Xaa represents Val, Leu, Ile, Thr, Asp, Glu, Asn, Gln, Lys,
; OTHER INFORMATION: Arg, Cys, Met, Phe, Tyr, Trp, His or Pro.
US-09-367-528A-1

Query Match          37.5%; Score 6; DB 9; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 GKDPV 14
Db      62 GKDPV 67

RESULT 13
US-09-367-528A-3
; Sequence 3, Application US/09367528A
; Publication No. US20010051359A1
; GENERAL INFORMATION:
; APPLICANT: TOYOTA JIDOSHA KABUSHIKI KAISHA
; TITLE OF INVENTION: Geranyl Diphosphate Synthetase Gene
; FILE REFERENCE: PH-586
; CURRENT APPLICATION NUMBER: US/09/367,528A
; CURRENT FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: JP97/346686
; PRIOR FILING DATE: 1997-12-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-367-528A-3

Query Match          37.5%; Score 6; DB 9; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 GKDPV 14
Db      62 GKDPV 67

RESULT 14
US-09-367-528A-5
; Sequence 5, Application US/09367528A
; Publication No. US20010051359A1
; GENERAL INFORMATION:
; APPLICANT: TOYOTA JIDOSHA KABUSHIKI KAISHA
; TITLE OF INVENTION: Geranyl Diphosphate Synthetase Gene
; FILE REFERENCE: PH-586
```

```
; CURRENT APPLICATION NUMBER: US/09/367,528A
; CURRENT FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: JP97/346686
; PRIOR FILING DATE: 1997-12-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-367-528A-5

Query Match          37.5%; Score 6; DB 9; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 GKDPV 14
Db      62 GKDPV 67

RESULT 15
US-10-462-698A-76
; Sequence 76, Application US/10462698A
; Publication No. US20040029239A1
; GENERAL INFORMATION:
; APPLICANT: TOYOTA JIDOSHA KABUSHIKI KAISHA
; TITLE OF INVENTION: A method of producing prenylalcohol
; FILE REFERENCE: PH-1412PCT
; CURRENT APPLICATION NUMBER: US/10/462,698A
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: JP2000-401701
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 76
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-10-462-698A-76

Query Match          37.5%; Score 6; DB 15; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 GKDPV 14
Db      62 GKDPV 67

Search completed: November 17, 2004, 13:21:36
Job time : 4.8405 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 09:38:53 ; Search time 0.830043 Seconds
(without alignments)
1854.686 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 16

Sequence: 1 FSWNCFEGKDPVIR 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 79: *
1: piri: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	9	56.2	162	2 S13195	ganglioside M2 act
2	9	56.2	193	2 I54178	ganglioside M2 act
3	9	56.2	200	2 S22411	ganglioside M2 act
4	8	50.0	20	2 S56005	lysosomal protein
5	8	50.0	193	2 S35613	ganglioside M2 act
6	7	43.8	391	1 F69459	probable vrpJ-ther
7	6	37.5	88	2 AG1370	hypothetical prote
8	6	37.5	126	2 B83265	hypothetical prote
9	6	37.5	201	2 G59096	hypothetical prote
10	6	37.5	212	2 AC1055	peptide methionine
11	6	37.5	243	2 B33329	cysteine-rich secr
12	6	37.5	249	2 E69340	cobalamin biosynth
13	6	37.5	260	2 T48846	creatine kinase (EC 3
14	6	37.5	297	2 JX0257	geranyltransferase
15	6	37.5	308	2 G97338	malonyl CoA-acyl c
16	6	37.5	329	2 C75400	UDP-N-acetylglucos
17	6	37.5	379	2 AI1391	hypothetical prote
18	6	37.5	425	2 AE2094	phosphoprotein pho
19	6	37.5	434	2 T04263	glutamate synthase
20	6	37.5	442	2 AI3075	glutamic protein [i
21	6	37.5	442	2 H98210	rtm protein (impor
22	6	37.5	530	2 AG3622	X-Pro aminopeptida
23	6	37.5	608	2 AI3325	conserved hypothet
24	6	37.5	833	2 A90575	sodium channel SCA
25	6	37.5	1993	2 T30902	25K acrosomal auto
26	5	31.2	20	2 A60802	chitinase (EC 3.2.
27	5	31.2	27	2 D44908	hypothetical prote
28	5	31.2	36	2 H82817	superoxide dismuta
29	5	31.2	44	2 S50088	

30 5 31.2 48 2 B85935 hypothetical prote
31 5 31.2 61 2 C81531 conserved hypothet
32 5 31.2 65 2 A41004 calgizzarin - chic
33 5 31.2 66 2 S10561 chlorophyll a/b-bi
34 5 31.2 70 2 T28640 Y4JM protein - Rni
35 5 31.2 72 2 T37089 probable IS elemen
36 5 31.2 72 2 AH32B1 hypothetical cytos
37 5 31.2 73 2 I47089 keratin type II -
38 5 31.2 73 2 D86616 CT849.1 hypothetic
39 5 31.2 73 2 C72008 CT849.1 hypothetic
40 5 31.2 90 2 C64013 hypothetical prote
41 5 31.2 91 2 AH0879 conserved hypothet
42 5 31.2 119 2 T26223 hypothetical prote
43 5 31.2 124 2 A97215 hypothetical prote
44 5 31.2 128 2 T30428 hypothetical prote
45 5 31.2 132 2 T49589 hypothetical prote

ALIGNMENTS

RESULT 1

S13195
ganglioside M2 activator protein - human
C;Species: Homo sapiens (man)
C;Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S13195
R;Fuerst, W.; Schubert, J.; Machleidt, W.; Meyer, H.E.; Sandhoff, K.
Eur. J. Biochem. 192, 709-714, 1990
A;Title: The complete amino-acid sequences of human ganglioside GM2 activator protein and
A;Reference number: S13195; MUID:91006165; PMID:2209618
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-162 <FUE>
A;Cross-references: UNIPROT:P17900

Query Match 56.2%; Score 9; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDPVIR 16
Db 10 EGKDPVIR 18
|||||

RESULT 2

I54178
ganglioside M2 activator protein precursor - human
C;Species: Homo sapiens (man)
C;Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
C;Accession: I54178; JQ1037; S05036; S22410; S17107
R;Xie, B.; Kennedy, J.L.; McInnes, B.; Auger, D.; Mahuran, D.
Genomics 14, 796-798, 1992
A;Title: Identification of a processed pseudogene related to the functional gene encoding
man chromosome 5.
A;Reference number: I54178; MUID:93052421; PMID:1427911
A;Accession: I54178
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-193 <RES>
A;Cross-references: UNIPROT:P17900; GB:I01439; NID:gl83358; PIDN:AAA52767.1; PID:gl83359
R;Xie, B.; McInnes, B.; Neote, K.; Lamhounwah, A.M.; Mahuran, D.
Biochem. Biophys. Res. Commun. 177, 1217-1223, 1991
A;Title: Isolation and expression of a full-length cDNA encoding the human GM2 activator
A;Reference number: JQ1037; MUID:91282768; PMID:2059210
A;Accession: JQ1037
A;Molecule type: mRNA
A;Residues: 1-18, 'A', 20-193 <XIE>
A;Cross-references: GB:M76477; NID:gl83356; PIDN:AAA35907.1; PID:gl83357
A;Experimental source: Hela cell
A;Note: 19-thr and 69-Met were also found
R;Schroeder, M.; Klima, H.; Nakano, T.; Kwon, H.; Quintern, L.E.; Gaertner, S.; Suzuki, T.

FEBS Lett. 251, 197-200, 1989
A;Title: Isolation of a cDNA encoding the human G(M2) activator protein.
A;Reference number: S05036; MUID:89325664; PMID:2753159
A;Accession: S05036
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 15-193 <SCH>
A;Cross-references: GB:X16087; NID:g31852; PIDN:CAA34215.1; PID:g31853
R;Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A;Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A;Reference number: S22410; MUID:92207171; PMID:1554364
A;Accession: S22410
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 43-142, 'VST', <NAG>
A;Cross-references: ENBL:X61094; NID:g31854; PIDN:CAA43407.1; PID:g31855
C;Comment: This protein transports GM2 ganglioside from the lysosomal membrane to hexosaminidase A.
C;Genetics:
A;Gene: GDB:GM2A
A;Cross-references: GDB:120000; OMIM:272750
A;Map position: 5q32-5q33
F;1-23/Domain: signal sequence #status predicted <SIG>
F;32-193/Product: GM2 ganglioside activator protein #status predicted <GM2>
F;32-193/Product: ganglioside M2 activator #status predicted <NAI>

Query Match 56.2%; Score 9; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.0066; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 8 EGKDPVAVR 16
DB 41 EGKDPVAVR 49

RESULT 3
S22411
A;Title: Isolation of a cDNA encoding the human (fragment) ganglioside M2 activator protein (clone pGAP2) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
C;Accession: S22411; S17108
R;Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A;Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A;Reference number: S22410; MUID:92207171; PMID:1554364
A;Accession: S22411
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-200 <NAG>
A;Cross-references: UNIPROT:P17900; EMBL:X61095; NID:g31856; PIDN:CAA43408.1; PID:g31857

Query Match 56.2%; Score 9; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 0.0068; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 8 EGKDPVAVR 16
DB 48 EGKDPVAVR 56

RESULT 4
S56005
A;Title: Isolation of a cDNA encoding the human (fragment) lysosomal protein 22X - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 07-May-1999
C;Accession: S56005
R;Kuwana, T.; Mullock, B.M.; Luzio, J.P.
Biochem. J. 308, 937-946, 1995
A;Title: Identification of a lysosomal protein causing lipid transfer, using a fluorescent probe.
A;Reference number: S56005; MUID:97104296; PMID:8948454
A;Accession: S56005
A;Status: preliminary
A;Molecule type: protein

A;Residues: 1-20 <KUW>

Query Match 50.0%; Score 8; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.013; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 8 EGKDPVAVR 15
DB 8 EGKDPVAVR 15

RESULT 5
S35613
A;Title: Isolation of a cDNA encoding the human (fragment) ganglioside M2 activator protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 07-May-1999
C;Accession: S35613
R;Bellachione, G.; Stirling, J.L.; Orlicchio, A.; Beccari, T.
Biochem. J. 294, 227-230, 1993
A;Title: Cloning and sequence analysis of a cDNA clone coding for the mouse G(M2) activator protein.
A;Reference number: S35613; MUID:93371367; PMID:7689829
A;Accession: S35613
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-193 <BEL>

Query Match 50.0%; Score 8; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.085; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 8 EGKDPVAVR 15
DB 41 EGKDPVAVR 48

RESULT 6
F69459
A;Title: Isolation of a cDNA encoding the human (fragment) probable vtpJ-therm 1 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: F69459
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
; Glodok, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Overbeek, R.; Cotton, M.D.; Spriggs, T.; Ariach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: F69459
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-391 <KLE>
A;Cross-references: UNIPROT:O28594; GB:AE000987; GB:AE000782; NID:g2689310; PIDN:AB88956
C;Superfamily: probable vtpJ-therm

Query Match 43.8%; Score 7; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.9; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 7 FEGKQPA 13
DB 276 FEGKQPA 282

RESULT 7
AG1970
A;Title: Isolation of a cDNA encoding the human (fragment) hypothetical protein asr1314 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AG1970
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AG1970
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-88 <KUR>
 A:Cross-references: UNIPROT:Q8YXA1; GB:BA000019; PIDN:BAW73271.1; PID:gl7130661; GSPDB:G
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: asr1314

Query Match 37.5%; Score 6; DB 2; Length 88;
 Best Local Similarity 100.0%; Pred. No. 7.2;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DPAVIR 16
 |||||
 Db 24 DPAVIR 29

RESULT 8
 B83265
 hypothetical protein PA3041 [imported] - Pseudomonas aeruginosa (strain PAO1)
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
 C:Accession: B83265
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
 A:Reference number: B82950; MUID:20437337; PMID:10984043
 A:Accession: B83265
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-126 <STO>
 A:Cross-references: UNIPROT:Q9HZG7; GB:AE004729; GB:AE004091; NID:g9949143; PIDN:AAG0642
 A:Experimental source: strain PAO1
 C:Genetics:
 A:Gene: PA3041
 C:Superfamily: membrane protein

Query Match 37.5%; Score 6; DB 2; Length 126;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ECKDPA 13
 |||||
 Db 3 ECKDPA 8

RESULT 9
 G59096
 hypothetical protein pX01-47 - Bacillus anthracis virulence plasmid pX01
 C:Species: Bacillus anthracis
 C>Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 09-Jul-2004
 C:Accession: G59096
 R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler
 J. Bacteriol. 181, 6509-6515, 1999
 A:Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harbori
 A:Reference number: A59091; MUID:99445483; PMID:10515943
 A:Accession: G59096
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-201 <OK1>
 A:Cross-references: UNIPROT:Q9X317; GB:AF065404; NID:g4894216; PIDN:AAD32351.1; PID:g489
 A:Experimental source: strain Sterne
 A>Note: similar to hypothetical, ORF4, plasmid pPOD2000; B. subtilis (U55043)
 C:Genetics:
 A:Gene: pX01-47
 A:Genome: plasmid

Query Match 37.5%; Score 6; DB 2; Length 201;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DPAVIR 16
 |||||
 Db 80 DPAVIR 85

RESULT 10
 AC1055
 Peptide methionine sulfoxide reductase [imported] - Salmonella enterica subsp. enterica s
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A>Note: this species has also been called Salmonella typhi
 C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AC1055
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 , S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serove
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AC1055
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-212 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD06888.1; PID:gl6505536; GSPDB:GN00176
 C:Genetics:
 A:Gene: msra
 C:Superfamily: peptide methionine sulfoxide reductase

Query Match 37.5%; Score 6; DB 2; Length 212;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DPAVIR 16
 |||||
 Db 102 DPAVIR 107

RESULT 11
 B33329
 cysteine-rich secretory protein 2 type I precursor - human
 N:Alternate names: testis-specific protein
 C:Species: Homo sapiens (man)
 C>Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 09-Jul-2004
 C:Accession: B33329; S88682
 R:Kasahara, M.; Gutknecht, J.; Brew, K.; Spurr, N.; Goodfellow, P.N.
 Genomics 5, 527-534, 1989
 A:Title: Cloning and mapping of a testis-specific gene with sequence similarity to a spe
 A:Reference number: A33329; MUID:90129048; PMID:2613236
 A:Accession: B33329
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-243 <XAS>
 R:Kraatzschmar, J.; Haendler, B.; Eberspaecher, U.; Roosterman, D.; Donner, P.; Schleunir
 Eur. J. Biochem. 236, 827-836, 1996
 A:Title: The human cysteine-rich secretory protein (CRISP) family. Primary structure and
 A:Reference number: S88681; MUID:96270732; PMID:8665901
 A:Accession: S88682
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-243 <KRA>
 A:Cross-references: EMBL:X95239; NID:gl262816; PIDN:CAA64526.1; PID:gl262817
 C:Genetics:
 A:Gene: GDB:TPX1
 A:Cross-references: GDB:120760; OMIM:187430
 A:Map position: 6p21-6qter
 C:Superfamily: cysteine-rich secretory protein 1
 F:i-20/Domain: signal sequence #status predicted <SIG>

F:21-243/Product: cysteine-rich secretory protein 2 type I #status predicted <MAT>

Query Match 37.5%; Score 6; DB 2; Length 243;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDP A 13
|||||
DB 20 EGKDP A 25

RESULT 12

E69340
cobalamin biosynthesis precorrin methylase (cbiG) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 16-Aug-2004
C:Accession: E69340
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: E69340
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-249 <KLE>
A:Cross-references: UNIPROT:Q29533; GB:AE001055; GB:AE000782; NID:g2689378; PIDN:AAB9051
C:Superfamily: Cobalamin biosynthesis protein CbiG

Query Match 37.5%; Score 6; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 KDP A V I 15
|||||
DB 85 KDP A V I 90

RESULT 13

T48846
creatininase (EC 3.5.2.10) [validated] - Pseudomonas sp.
N:Alternate names: creatinine amidohydrolase
C:Species: Pseudomonas sp.
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C:Accession: T48846
R:Yamamoto, K.; Oka, M.; Kikuchi, T.; Emi, S.
Biosci. Biotechnol. Biochem. 59, 1331-1332, 1995
A:Title: Cloning of the creatinine amidohydrolase gene from pseudomonas sp. PS-7.
A:Reference number: Z24553; MUID:95400012; PMID:7670196
A:Accession: T48846
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-260 <YAM>
A:Cross-references: UNIPROT:Q52548; EMBL:D45424; NID:g662293; PIDN:BAA08265.1; PID:g6622
A:Experimental source: strain PS-7
C:Function:
A:Description: EC 3.5.2.10 [validated, MUID:95400012]
C:Keywords: hydrolase

Query Match 37.5%; Score 6; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 KDP A V I 15
|||||
DB 158 KDP A V I 163

RESULT 14

JX0257

geranyltransterase (EC 2.5.1.10) - Bacillus stearothermophilus
N:Alternate names: farnesyl-diphosphate synthase

C:Species: Bacillus stearothermophilus
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: JX0257
R:Koyama, T.; Obata, S.; Osabe, M.; Takeshita, A.; Yokoyama, K.; Uchida, M.; Nishino, T.
J. Biochem. 113, 355-363, 1993
A:Title: Thermostable farnesyl diphosphate synthase of Bacillus stearothermophilus: molec
A:Reference number: JX0257; MUID:93252758; PMID:8486607
A:Accession: JX0257
A:Molecule type: DNA
A:Residues: 1-297 <KOY>
A:Cross-references: UNIPROT:Q08291; GB:D13293; NID:g391609; PIDN:BAA02551.1; PID:g391610
C:Genetics:
A:Start codon: GTG
C:Superfamily: Geranyltransterase
C:Keywords: transferase

Query Match 37.5%; Score 6; DB 2; Length 297;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GKDP A V 14
|||||
DB 62 GKDP A V 67

RESULT 15

G97338
malonyl CoA-acyl carrier protein transacylase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: G97338
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97338
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-308 <KUR>
A:Cross-references: UNIPROT:Q97DA5; GB:AE001437; PIDN:AAK81498.1; PID:g15026671; GSPDB:G
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3575
C:Superfamily: [acyl-carrier-protein] S-malonyltransferase; [acyl-carrier-protein] S-mal

Query Match 37.5%; Score 6; DB 2; Length 308;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 CFE G K D 11
|||||
DB 48 CFE G K D 53

Search completed: November 17, 2004, 10:47:51
Job time : 2.83004 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 09:35:07 ; Search time 4.63337 Seconds
(without alignments)
1986.889 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 16

Sequence: 1 FSWDNCFEGKDAVIR 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Uniprot_02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	56.2	47	Q8I028	Q8I028 macaca fasc
2	9	56.2	189	Q6L8L5	Q6L8L5 homo sapien
3	9	56.2	189	CAAA3394	CAAA3394 homo sapi
4	9	56.2	190	Q8HXX6	Q8HXX6 macaca fasc
5	9	56.2	193	1 SAP3 HUMAN	P17900 homo sapien
6	8	50.0	20	Q9QUW2	Q9QUW2 rattus sp.
7	8	50.0	193	1 SAP3 MOUSE	Q6648 mus musculu
8	8	50.0	199	Q6IN37	Q6IN37 rattus norv
9	8	50.0	199	Q8CJH4	Q8CJH4 rattus norv
10	7	43.8	103	Q14427	Q14427 homo sapien
11	7	43.8	391	Q2B594	Q2B594 archaeglob
12	7	43.8	592	Q7VCL2	Q7VCL2 prochloroco
13	6	37.5	88	Q8YXA1	Q8YXA1 anabaena sp
14	6	37.5	120	Q8TQY1	Q8TQY1 methanosarc
15	6	37.5	126	Q9HZG7	Q9HZG7 pseudomonas
16	6	37.5	129	Q7YW32	Q7YW32 paramecium
17	6	37.5	129	Q7YW33	Q7YW33 paramecium
18	6	37.5	129	Q7YW34	Q7YW34 paramecium
19	6	37.5	129	Q7YW36	Q7YW36 paramecium
20	6	37.5	146	Q6Q7X4	Q6Q7X4 felis silve
21	6	37.5	146	Q6Q7X5	Q6Q7X5 felis silve
22	6	37.5	149	Q6SQE8	Q6SQE8 meiotherm
23	6	37.5	149	Q6Q7X5	Q6Q7X5 felis silve
24	6	37.5	151	Q9N0J2	Q9N0J2 bos taurus
25	6	37.5	151	Q9N0J2	Q9N0J2 bos taurus
26	6	37.5	151	Q9N0J2	Q9N0J2 bos taurus
27	6	37.5	160	Q6I2W5	Q6I2W5 picophilus
28	6	37.5	166	Q8ETE8	Q8ETE8 homo sapien
29	6	37.5	170	1 P25B HUMAN	P59282 homo sapien
30	6	37.5	188	Q9M7W7	Q9M7W7 arabidopsis
31	6	37.5	196	Q8TVA1	Q8TVA1 methanopyru

32 6 37.5 201 2 Q7CMH7
33 6 37.5 201 2 Q9X317
34 6 37.5 201 2 AAT28810
35 6 37.5 212 1 MSRA_SALTI
36 6 37.5 243 1 CRS2_HUMAN
37 6 37.5 243 2 AAP41200
38 6 37.5 243 2 AAP41114
39 6 37.5 244 2 Q8HX97
40 6 37.5 249 2 O29533
41 6 37.5 260 2 P83772
42 6 37.5 260 2 Q52548
43 6 37.5 261 2 Q6M141
44 6 37.5 261 2 CAE78141
45 6 37.5 278 2 Q7Z7B2

ALIGNMENTS

RESULT 1
Q8I028 PRELIMINARY; PRT; 47 AA.
AC Q8I028;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Ganglioside GM2 activator (Fragment).
GN Name=gm2a;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RA Kusuda J., Osada N., Hashimoto K.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB083332; BAC20643.1; -.
FT NON TER 1 1
FT NON TER 47 47
SQ SEQUENCE 47 AA; 4893 MW; C5F5537F3A029FFB CRC64;

Query Match 56.2%; Score 9; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EKGDAVIR 16
Db 11 EKGDAVIR 19

RESULT 2
Q6L8L5 PRELIMINARY; PRT; 189 AA.
ID Q6L8L5
AC Q6L8L5;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE GM2 activator protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=94008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
human GM2 activator protein."
RL FEBS Lett. 289:260-264(1991).
RN [2]
RP SEQUENCE FROM N.A.

RX MEDLINE=93277527; PubMed=8503891;
RA Klima H., Klein A., Van Echten G., Schwarzmam G., Suzuki K.,
RA Sandhoff K.;
RT "Over-expression of a functionally active human Gm2-activator protein
in escherichia coli.";
RL Biochem. J. 292:571-576(1993).
DR EMBL; X62078; CAA43994.1; -;
DR InterPro; IPR003172; E1_DerP2_DerF2.
DR SMART; SMO0737; ML; 1.
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match 56.2%; Score 9; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 EGKDPVAVIR 16
| | | | | | | |
Db 37 EGKDPVAVIR 45

RESULT 3
CAA43994 PRELIMINARY; PRT; 189 AA.
ID CAA43994;
AC CAA43994;
DT 02-MAR-2004 (TRENBLrel. 27, Created)
DT 02-MAR-2004 (TRENBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TRENBLrel. 27, Last annotation update)
DE Gm2 activator protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
human Gm2 activator protein.";
RL FEBS Lett. 289:260-264(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277527; PubMed=8503891;
RA Klima H., Klein A., Van Echten G., Schwarzmam G., Suzuki K.,
RA Sandhoff K.;
RT "Over-expression of a functionally active human Gm2-activator protein
in escherichia coli.";
RL Biochem. J. 292:571-576(1993).
DR EMBL; X62078; CAA43994.1; -;
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match 56.2%; Score 9; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 EGKDPVAVIR 16
| | | | | | | |
Db 37 EGKDPVAVIR 45

RESULT 4
Q8HXX6 PRELIMINARY; PRT; 190 AA.
ID Q8HXX6;
AC Q8HXX6;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Ganglioside Gm2 activator.
GN Name=gsm2a;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.

OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain cerebellum cortex;
RA Kusuda J., Osada N., Hida M., Sugano S., Hashimoto K.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB083313; BAC20592.1; -;
DR HSSP; P17900; LG13
DR InterPro; IPR003172; E1_DerP2_DerF2.
DR SMART; SMO0737; ML; 1.
SQ SEQUENCE 190 AA; 20494 MW; 9F9582BBE75715C3 CRC64;

Query Match 56.2%; Score 9; DB 2; Length 190;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 EGKDPVAVIR 16
| | | | | | | |
Db 38 EGKDPVAVIR 46

RESULT 5
SAP3 HUMAN STANDARD; PRT; 193 AA.
ID SAP3 HUMAN
AC P17900; Q14426; Q14428;
DT 01-NOV-1990 (Rel. 15, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Ganglioside Gm2 activator precursor (GM2-AP) (Cerebroside sulfate
DE activator protein) (Shingolipid activator protein 3) (SAP-3).
GN Name=GM2A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RX MEDLINE=91282768; PubMed=2059210;
RA Xie B., McInnes B., Neote K., Lamhonwah A.-M., Mahuran D.;
RT "Isolation and expression of a full-length cDNA encoding the human G-
M2 activator protein.";
RL Biochem. Biophys. Res. Commun. 177:1217-1223(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
human Gm2 activator protein.";
RL FEBS Lett. 289:260-264(1991).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RC TISSUE=Placenta;
RX MEDLINE=92207171; PubMed=1554364;
RA Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
RT "Evidence for two cDNAs encoding human GM2-activator protein.";
RL Biochem. J. 282:807-813(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93052421; PubMed=1427911;
RA Xie B., Kennedy J.L., McInnes B., Auger D., Mahuran D.J.;
RT "Identification of a processed pseudogene related to the functional
gene encoding the GM2 activator protein: localization of the
pseudogene to human chromosome 3 and the functional gene to human
chromosome 5.";
RL Genomics 14:796-798(1992).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=99294584; PubMed=10364519;
RA Chen B., Rigat B., Curry C., Mahuran D.J.;
RT "Structure of the GM2A gene: identification of an exon 2 nonsense
mutation and a naturally occurring transcript with an in-frame
deletion of exon 2.";

FT STRAND 107 108
 FT TURN 109 110
 FT HELIX 111 118
 FT TURN 121 122

Query Match 56.2%; Score 9; DB 1; Length 193;
 Best Local Similarity 100.0%; Pred. No. 0.023;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 EGKDPVAVR 16

Db 41 EGKDPVAVR 49

RESULT 6

Q9QW2

ID Q9QW2 PRELIMINARY; PRT; 20 AA.

AC Q9QW2;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE 22 kDa INTRALYSOSOMAL lipid-transfer protein (Fragment).

OS Rattus sp.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10118;

RN [1]

RP SEQUENCE.

RX MEDLINE=97104296; PubMed=948454;

RA Kuwana T., Mullock B.M., Iuzio J.P.;

RT "Identification of a lysosomal protein causing lipid transfer, using a

RT fluorescence assay designed to monitor membrane fusion between rat

RT liver endosomes and lysosomes.";

RL Biochem. J. 308:937-946(1995).

DR HSPF; P17900; IG13.

SQ SEQUENCE 20 AA; 2211 MW; 3A160591EA5542C5 CRC64;

Query Match 50.0%; Score 8; DB 2; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.042;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 EGKDPVAVI 15

Db 8 EGKDPVAVI 15

RESULT 7

SAP3 MOUSE

ID SAP3 MOUSE STANDARD; PRT; 193 AA.

AC Q60648; Q61610; Q61819;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 05-JUN-2004 (Rel. 44, Last annotation update)

DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate

DE activator protein) (Shingolipid activator protein 3) (SAP-3).

GN Name=Gm2a;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;

RX MEDLINE=95229165; PubMed=7713516;

RA Yamanaka S., Johnson O.N., Lyu M.S., Kozak C.A., Proia R.L.;

RT "The mouse gene encoding the GM2 activator protein (Gm2a): cDNA

RT sequence, expression, and chromosome mapping.";

RL Genomics 24:601-604(1994).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=93371367; PubMed=7689829;

RA Bellachioma G., Stirling J.L., Orlacchio A., Beccari T.;

RT "Cloning and sequence analysis of a cDNA clone coding for the mouse

GM2 activator protein.";
 Biochem. J. 294:227-230(1993).

[3]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6 X CBA;

RX MEDLINE=97224573; PubMed=9060405;

RA Bertoni C., Appolloni M.G., Stirling J.L., Li S.C., Li Y.T.,

RA Orlacchio A., Beccari T.;

RT "Structural organization and expression of the gene for the mouse GM2

RT activator protein.";

RL Mamm. Genome 8:90-93(1997).

[4]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Mammary gland;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: Binds gangliosides and stimulates ganglioside GM2

CC degradation. It stimulates only the breakdown of ganglioside GM2

CC and glycolipid GA2 by beta-hexosaminidase A. It extracts single

CC GM2 molecules from membranes and presents them in soluble form to

CC beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and

CC conversion to GM3.

CC -!- SUBCELLULAR LOCATION: Lysosomal.

CC -!- TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney and

CC testis.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; U09816; AAA21543.1; -

EMBL; L19526; AAA61929.1; -

EMBL; U34359; AAB06275.1; ALT SEQ.

EMBL; U34356; AAB06275.1; JOINED.

EMBL; U34357; AAB06275.1; JOINED.

EMBL; U34358; AAB06275.1; JOINED.

EMBL; BC004651; AA04651.1; -

HSPF; P17900; IG13.

MGD; MGI:95762; Gm2a.

Glycoprotein; Lysosome; Signal; Sphingolipid metabolism.

By similarity.

CHAIN 1 31

FT 32 193 Ganglioside GM2 activator.

FT 39 183 By similarity.

FT 99 106 By similarity.

FT 112 138 By similarity.

FT 125 136 By similarity.

FT 151 151 N-linked (GlcNAc...) (Potential).

FT 53 53 I -> T (in Ref. 1).

SQ SEQUENCE 193 AA; 20824 MW; 59CC4ABE56FA1FC7 CRC64;

```
Query Match          50.0%; Score 8; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDPDAVI 15
DB 41 EGKDPDAVI 48

RESULT 8
Q6IN37          PRELIMINARY; PRT; 199 AA.
AC Q6IN37;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE GM2 ganglioside activator protein.
GN Name=Gm2a;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Miwa N., Okada T., Nakamura S.;
RA Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB051391; BAC24018.1; -.
DR HSPL; P17900; IG13.
DR InterPro; IPR003172; EI_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 199 AA; 21521 MW; D585203FDFABF507 CRC64;

Query Match          50.0%; Score 8; DB 2; Length 199;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDPDAVI 15
DB 47 EGKDPDAVI 54

RESULT 10
Q14427          PRELIMINARY; PRT; 103 AA.
AC Q14427;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GM2-activator protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Placenta;
RX MEDLINE=92207171; PubMed=1554364;
RA Nagakajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
RT "Evidence for two cDNAs encoding human GM2-activator protein.";
RL Biochem. J. 282:807-813 (1992).
DR EMBL; X61094; CAA43407.1; -.
FT NON TER 1
SQ SEQUENCE 103 AA; 11137 MW; 434BFC334E976F91 CRC64;

Query Match          43.8%; Score 7; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 KDPDAVIR 16
DB 1 KDPDAVIR 7

RESULT 11
O28594          PRELIMINARY; PRT; 391 AA.
AC O28594;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE VtpJ-therm, putative.
GN OrderedLocusNames=AF1679;
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RC MEDLINE=98049343; PubMed=9389475; DOI=10.1038/37052;
RX
```

```

RA Klenk H.-P., Clayton R.A., Tomb J.-P., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M.L., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.R.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.J.,
RA Peterson S.N., Reich C.I., McNeil L.K., Badger J.H., Glodek A.,
RA Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L.A.,
RA Utterback T.R., Cotton M.D., Spriggs T., Artiach P., Kaine B.P.,
RA Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,
RA Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,
RA Woese C.R., Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AF000987; AAB89567.1; -.
DR PIR; F69459; F69459.
DR TIGR; AF1679; -.
KW Complete proteome.
SQ SEQUENCE 391 AA; 44328 MW; 49136BB0786DE3F4 CRC64;

Query Match 43.8%; Score 7; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 FEGKDP A 13
Db 276 FEGKDP A 282

RESULT 12
Q7VCL2 PRELIMINARY; PRT; 592 AA.
AC Q7VCL2;
DT 01-OCT-2003 (TremBLrel. 25, Created)
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Glycosidase.
GN Name=amyA; OrderedLocNames=Pro0728;
OS Prochlorococcus marinus.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=1219;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SARG / CCMP 1375 / SS120;
RX MEDLINE=22810154; PubMed=12917486;
RA Dufresne A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,
RA Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,
RA Makarova K.S., Ostrowski M., Ostas S., Robert C., Rogozin I.B.,
RA Scaulan D.J., Tandeau de Marsac N., Weissenbach J., Wincker P.,
RA Wolf Y.I., Hess W.R.;
RT "Genome sequence of the cyanobacterium Prochlorococcus marinus SS120,
RT a nearly minimal oxyphototrophic genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025(2003).
DR EMBL; AB017163; AAP99772.1; -.
DR GO; GO:0004556; F:alpha-amyase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR006047; Alpha_amiyl_cat.
DR Pfam; PF00128; Alpha-amyase; 1.
DR PROSITE; PS00141; ASP PROTEASE; UNKNOWN_1.
KW Complete proteome; Glycosidase.
SQ SEQUENCE 592 AA; 67693 MW; 592C6525433A0143 CRC64;

Query Match 43.8%; Score 7; DB 2; Length 592;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 DNFCEGK 10
Db 51 DNFCEGK 57

```

```

RESULT 13
Q8YXAL PRELIMINARY; PRT; 88 AA.
AC Q8YXAL;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Asr1314 protein.
GN OrderedLocNames=asr1314;
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpou S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AF003585; BAB73271.1; -.
DR PIR; AG1970; AG1970.
KW Complete proteome.
SQ SEQUENCE 88 AA; 9728 MW; 84C021A661DA41AC CRC64;

Query Match 37.5%; Score 6; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DPAVIR 16
Db 24 DPAVIR 29

RESULT 14
Q8TQY1 PRELIMINARY; PRT; 120 AA.
AC Q8TQY1;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Predicted protein.
GN OrderedLocNames=MA1406;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nushbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Anoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario B.C.,
RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010810; AAM04821.1; -.
KW Complete proteome.
SQ SEQUENCE 120 AA; 13724 MW; 8472323E39E440F0 CRC64;

```

Query Match 37.5%; Score 6; DB 2; Length 120;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 KDPABI 15
 |||||
 DB 90 KDPABI 95

RESULT 15

Q9HZG7 PRELIMINARY; PRT; 126 AA.
 AC Q9HZG7
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocusNames=PA3041;
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RA MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964 (2000).
 DR EMBL; AE004729; AAC06429.1; -.
 DR PIR; B83265; B83265.
 DR InterPro; IPR011309; UCP009726.
 DR PIRSF; PIRSF009726; UCP009726; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 126 AA; 13919 MW; 2DE48A3F7FEC34E9 CRC64;

Query Match 37.5%; Score 6; DB 2; Length 126;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 EGKDP 13
 |||||
 DB 3 EGKDP 8

Search completed: November 17, 2004, 10:45:34
 Job time : 6.63337 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 17:56:16 ; Search time 651.378 Seconds
(without alignments)
1161.592 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 94

Sequence: 1 FSWDNCFEGKDPVIR 16

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPTO.spool.p/US10030937/runat_16112004_153013_2913/app.query.fasta_1.789
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosu62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937@cgn2.1.7771@runat_16112004_153013_2913 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	90.4	257	9 AB083332	AB083332 Macaca fa
2	85	90.4	352	11 G05651	G05651 human STS w
3	85	90.4	821	9 HSGM2AP	X16087 Human mRNA
4	85	90.4	950	9 AB083313	AB083313 Macaca fa

5	90.4	953	9 HUMGM2	M76477 Human G-M2
6	90.4	1043	9 HUMGM2A	I01439 Human GM2-a
7	90.4	1045	6 CQ728078	CQ728078 Sequence
8	90.4	1047	9 HSGM2A2	AF124718 Homo sapi
9	90.4	1093	9 HSGM2APB	X61095 H.sapiens R
10	90.4	2413	9 BC009273	BC009273 Homo sapi
11	90.4	2436	6 AX330938	AX330938 Sequence
12	90.4	2436	9 HSGM2AFT	X62078 H.sapiens m
13	90.4	107320	2 AC011391	AC011391 Homo sapi
14	90.4	120584	9 AC011342	AC011342 Homo sapi
15	90.4	151712	9 AC008385	AC008385 Homo sapi
16	87.2	224	10 WMGM2AP2	U34357 Mus musculu
17	87.2	600	10 AB051391	AB051391 Rattus no
18	87.2	1113	10 MUSGM2ACT	L19526 Mouse GM2 a
19	87.2	1983	6 E12286	E12286 cDNA encodi
20	87.2	1983	6 AX827433	AX827433 Sequence
21	87.2	2003	10 MMU09816	U09816 Mus musculu
22	87.2	2024	10 BC004651	BC004651 Mus muscu
23	87.2	2028	10 BC072474	BC072474 Rattus no
24	87.2	39262	10 AL772357	AL772357 Mouse DNA
25	87.2	189516	2 AC128065	AC128065 Rattus no
26	87.2	242756	2 AC093965	AC093965 Rattus no
27	87.2	249474	2 AC136421	AC136421 Rattus no
28	85.1	321	6 CQ736066	CQ736066 Sequence
29	80	529	9 AB051291	AB051291 Homo sapi
30	85.1	564	9 HUMGM2APC	L01440 Human GM2A
31	85.1	111861	9 AC069435	AC069435 Homo sapi
32	58	928	5 BX950406	BX950406 Gallus ga
33	58	100575	9 HSDJ81F6	AL049762 Human DNA
34	56	134973	9 AP000818	AP000818 Homo sapi
35	56	153632	9 AC090938	AC090938 Homo sapi
36	56	160492	9 AP001884	AP001884 Homo sapi
37	56	177779	2 AC116870	AC116870 Mus muscu
38	56	219836	9 AY366501	AY366501 Homo sapi
39	55	139635	9 AC013290	AC013290 Homo sapi
40	55	145124	2 AC015672	AC015672 Homo sapi
41	55	148919	2 AC020680	AC020680 Homo sapi
42	55	159505	9 AC105042	AC105042 Homo sapi
43	55	172626	9 AC105101	AC105101 Homo sapi
44	55	175384	9 AC103774	AC103774 Homo sapi
45	55	176439	2 AC069131	AC069131 Homo sapi

ALIGNMENTS

AB083332 257 bp DNA linear PRI 09-OCT-2002
Macaca fascicularis gm2a gene for ganglioside GM2 activator,
partial cds, exon 2.
AB083332
AB083332.1 GI:23616929
Macaca fascicularis (crab-eating macaque)
Macaca fascicularis
Macaca fascicularis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
Cercopithecinae; Macaca.

REFERENCE

1
Kusuda,J., Osada,N. and Hashimoto,K.
Cloning of partial genomic sequence of cynomolgus monkey GM2A gene
Unpublished
2 (bases 1 to 257)
Kusuda,J.
Direct Submission
Submitted (08-APR-2002) Jun Kusuda, National Institute of
Infectious Diseases, Division of Genetic Resources; Toyama-cho,
Shinjuku, Tokyo 162-8640, Japan [E-mail:jkusuda@nih.go.jp.
URL:http://www.nih.go.jp, Tel:81-3-5285-1111(ex.2122),
Fax:81-3-5285-1181)

FEATURES

source

Location/Qualifiers
1..257
/organism="Macaca fascicularis"

```

/mol_type="genomic DNA"
/db_xref="taxon:9541"
/chromosome="5"
/map="Sq31.3-q33.1"
/note="PCR product"
116..257
/gene="gm2a"
<116..257
/gene="gm2a"
/codon_start=1
/product="ganglioside GM2 activator"
/protein_id="PAC20643.1"
/db_xref="GI:23616930"
/translation="IGSFSWDCDEGKDPVIRSLTLEPDPILIPGNVTVSVGSTSV
PLS"
116..257
/gene="gm2a"
/number=2

exon

ORIGIN
Alignment Scores:
Pred. No.: 1.21e-06 Length: 257
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-68 (1-16) x AB083332 (1-257)

QY 1 PhseSTpApcApcCysPheGluGlyAspProAlaValIleArg 16
DB 125 TTCTCTGGGATAACTGTGATGAAGAAAGACCCCTCGGTGATCAGA 172

RESULT 2
LOCUS G05651 352 bp DNA linear STS 19-OCT-1995
DEFINITION human STS WI-6085, sequence tagged site.
ACCESSION G05651
VERSION G05651.1 GI:858896
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 352)
AUTHORS Hudson,T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
JOURNAL Mapped ESTs
COMMENT Unpublished (1995)
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: AGCTCAGTAGCTTTTCCTGGG
Primer B: CCTTCTCCAAACTAAATCCACC
STS size: 177
PCR Profile:
Presoak:
Denaturing:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 nM

```

Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3

Prepared with primer pairs derived from T28061 -- dbEST.

FEATURES

Location/Qualifiers

1..352
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="749_D_6; 850_E_3; 958_C_10; 775_A_(6,7)"
STS
84..260
primer_bind
84..104
primer_bind
complement(238..260)

ORIGIN

Alignment Scores:
Pred. No.: 1.69e-06 Length: 352
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 11 Gaps: 0

US-10-030-937-68 (1-16) x G05651 (1-352)

QY 1 PhseSTpApcApcCysPheGluGlyAspProAlaValIleArg 16
DB 95 TTCTCTGGGATAACTGTGATGAAGAAAGACCCCTCGGTGATCAGA 142

RESULT 3

LOCUS HSGM2AP 821 bp mRNA linear PRI 19-JUL-1995
DEFINITION Human mRNA for G(M2) activator protein.
ACCESSION X16087
VERSION X16087.1 GI:31852
KEYWORDS G(M2) activator protein; G(M2) gangliosidosis.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 821)
AUTHORS Gartner,M., Klima,H., Nakano,T., Kwon,H., Quintern,L.E.,
Schroder,M., Suzuki,K. and Sandhoff,K.
TITLE Isolation of a cDNA encoding the human GM2 activator protein
JOURNAL FEBS Lett. 251 (1-2), 197-200 (1989)
MEDLINE 89325664
PUBMED 2753159
REFERENCE 2 (bases 1 to 821)
AUTHORS Klima,H., Klein,A., van Echten,G., Schwarzmann,G., Suzuki,K. and
Sandhoff,K.
TITLE Over-expression of a functionally active human GM2-activator
JOURNAL Biochem. J. 292 (Pt 2), 571-576 (1993)
MEDLINE 93277527
PUBMED 8503891
COMMENT Data kindly reviewed (23-JAN-1991) by Sandhoff K.
FEATURES
Location/Qualifiers
1..821
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="pGAP1"
/cell_type="fibroblast"
1..540
/codon_start=1
/product="G(M2) activator protein"
/protein_id="CAA34215.1"

CDS


```

/db_xref="GI:31853"
/db_xref="GOA:P17900"
/db_xref="Swiss-Prot:P17900"
/translation="LLATPAQHLKPKSLSFSWDCGKDPVIRSLTLLEPPI
VPGNVTLSVVGSTVPLSPKVLVLEKEVAGLWIKIPCTDIIGSTPEHFCVDLD
MLPTGPECPPEPLRTYGLPCHCPKEGTSLPKSEFVVPDLEPLSWLTITGNRIEVL
SSSGKRLGCIKIAASLKG1"
<1..51
sig_peptide 52..537
mat_peptide /product="G(M2) activator protein"
/evidence=experimental

ORIGIN
Alignment Scores:
Pred. No.: 4,21e-06 Length: 821
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-68 (1-16) x HSGM2AP (1-821)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 58 TTCTCTGGGATAACTGTGATGAAGGAGGACCTCGCGTGATCAGA 105

RESULT 4
AB083313
LOCUS Macaca fascicularis gm2a mRNA for ganglioside GM2 activator,
DEFINITION complete cds.
ACCESSION AB083313
VERSION AB083313.1 GI:23574732
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Macaca fascicularis (crab-eating macaque)
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Macaca.

REFERENCE
1 Kusuda,J., Osada,N., Hida,M., Sugano,S. and Hashimoto,K.
Isolation and characterization of cDNA for macaque neurological
disease genes
Unpublished
2 (bases 1 to 950)
Kusuda,J.
Direct Submission
Submitted (08-APR-2002) Jun Kusuda, National Institute of
Infectious Diseases, Division of Genetic Resources; Toyama,
Shinjuku, Tokyo 1628640, Japan (E-mail:jkusuda@nih.go.jp,
URL:http://www.nih.go.jp, Tel:81-3-5285-1111(ex.2122),
Fax:81-3-5285-1181)
Location/Qualifiers
1..950
/organism="Macaca fascicularis"
/mol_type="mRNA"
/db_xref="taxon:9541"
/clone="QCCB-17591"
/tissue_type="brain cerebellum cortex"
/clone_lib="macaque brain library QCCB"
/note="vector:TOP10"
1..950
/gene="gm2a"
66..638
/gene="gm2a"
/codon_start=1
/product="ganglioside GM2 activator"
/protein_id="BAC20592.1"
/db_xref="GI:23574733"
/translation="MQSLMQAPLLIALGLLFAAPQAHLKPKSLSFSWDCGKDPV
IRSLTLEPPIIPGNVTLSVVGSTVPLSPKVLVLEKEVAGLWIKIPCTDIIGS
CTFEDSCDVLMDLITPGECPEPLATYGLPCHCPKEGTSLPKSEFVVPDLEPLSWL

```

```

TTGNRIEISILSNRGRKLGCIKIAASLKG1"

ORIGIN
Alignment Scores:
Pred. No.: 4,92e-06 Length: 950
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-68 (1-16) x AB083313 (1-950)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 156 TTCTCTGGGATAACTGTGATGAAGGAGGACCTCGCGTGATCAGA 203

RESULT 5
HUMGM2
LOCUS Human G-M2 activator protein mRNA, complete cds.
DEFINITION Human G-M2 activator protein mRNA, complete cds.
ACCESSION M76477
VERSION M76477.1 GI:183356
KEYWORDS G-M2 activator protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 953)
Xie,B., McInnes,B., Neote,K., Lamhonwah,A.M. and Mahuran,D.
Isolation and expression of a full-length cDNA encoding the human
GM2 activator protein
JOURNAL Biochem. Biophys. Res. Commun. 177 (3), 1217-1223 (1991)
MEDLINE 91282768
PubMed 2059210
COMMENT Original source text: Homo sapiens mRNA.
FEATURES
Location/Qualifiers
1..953
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
91..672
/codon_start=1
/product="G-M2 activator protein"
/protein_id="AAA35907.1"
/db_xref="GI:183357"
/translation="MQSLMQAPLLIALGLLFAAPQAHLKPKSLSFSWDCGKDPV
IRSLTLEPPIIPGNVTLSVVGSTVPLSPKVLVLEKEVAGLWIKIPCTDY
IGSCTFEHPFCDVLMDLITPGECPEPLRTYGLPCHCPKEGTSLPKSEFVVPDLEPL
SWLTITGNRIEISILSNRGRKLGCIKIAASLKG1"
91..159
sig_peptide
mat_peptide 160..669
/product="G-M2 activator protein"

ORIGIN
Alignment Scores:
Pred. No.: 4,94e-06 Length: 953
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-68 (1-16) x HUMGM2 (1-953)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 190 TTCTCTGGGATAACTGTGATGAAGGAGGACCTCGCGTGATCAGA 237

RESULT 6
HUMGM2A
LOCUS Human GM2-activator protein (GM2A) mRNA, complete cds.
DEFINITION Human GM2-activator protein (GM2A) mRNA, complete cds.

```

```

1..950
/organism="Macaca fascicularis"
/mol_type="mRNA"
/db_xref="taxon:9541"
/clone="QCCB-17591"
/tissue_type="brain cerebellum cortex"
/clone_lib="macaque brain library QCCB"
/note="vector:TOP10"
1..950
/gene="gm2a"
66..638
/gene="gm2a"
/codon_start=1
/product="ganglioside GM2 activator"
/protein_id="BAC20592.1"
/db_xref="GI:23574733"
/translation="MQSLMQAPLLIALGLLFAAPQAHLKPKSLSFSWDCGKDPV
IRSLTLEPPIIPGNVTLSVVGSTVPLSPKVLVLEKEVAGLWIKIPCTDIIGS
CTFEDSCDVLMDLITPGECPEPLATYGLPCHCPKEGTSLPKSEFVVPDLEPLSWL

```

```

1..953
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
91..672
/codon_start=1
/product="G-M2 activator protein"
/protein_id="AAA35907.1"
/db_xref="GI:183357"
/translation="MQSLMQAPLLIALGLLFAAPQAHLKPKSLSFSWDCGKDPV
IRSLTLEPPIIPGNVTLSVVGSTVPLSPKVLVLEKEVAGLWIKIPCTDY
IGSCTFEHPFCDVLMDLITPGECPEPLRTYGLPCHCPKEGTSLPKSEFVVPDLEPL
SWLTITGNRIEISILSNRGRKLGCIKIAASLKG1"
91..159
sig_peptide
mat_peptide 160..669
/product="G-M2 activator protein"

ORIGIN
Alignment Scores:
Pred. No.: 4,94e-06 Length: 953
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-68 (1-16) x HUMGM2 (1-953)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 190 TTCTCTGGGATAACTGTGATGAAGGAGGACCTCGCGTGATCAGA 237

RESULT 6
HUMGM2A
LOCUS Human GM2-activator protein (GM2A) mRNA, complete cds.
DEFINITION Human GM2-activator protein (GM2A) mRNA, complete cds.

```

```
ACCESSION L01439
VERSION L01439.1 GI:183358
KEYWORDS GM2 activator protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1043)
AUTHORS Xie, B., Kennedy, J.L., McInnes, B., Auger, D. and Mahuran, D.
TITLE Identification of a processed pseudogene related to the functional
gene encoding the GM2 activator protein: localization of the
pseudogene to human chromosome 3 and the functional gene to human
chromosome 5
JOURNAL Genomics 14 (3), 796-798 (1992)
MEDLINE 93052421
PUBMED 1427911
COMMENT Original source text: Homo sapiens cDNA to mRNA.
FEATURES
    source
        1..1043
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /map="5"
        1..1043
            /gene="GM2A"
            /protein_id="AA052767.1"
            /db_xref="GI:183359"
            /translation="MQSLMQAPLLIALGLLLATPAQAHLKPQSLSPSWNCDGKDD
            PAVRSTLTLEDPGVNPGNTLVSVGTSVPLSPKVDLVLEKAVGLWIKIPCTDY
            IGSTCFHFCDVLDMLPTGPCEPLRTYGLPCHCPKFGTSLPKSEFVVPDLELP
            SWLTGNVRIESLVSSSGKRLGCIKIASLKG1"
    gene
        1..1043
            /gene="GM2A"
    CDS
        91..672
            /codon_start=1
            /product="GM2-activator protein"
            /protein_id="AA052767.1"
            /db_xref="GI:183359"
            /db_xref="GDB:1000-120-000"
            /translation="MQSLMQAPLLIALGLLLATPAQAHLKPQSLSPSWNCDGKDD
            PAVRSTLTLEDPGVNPGNTLVSVGTSVPLSPKVDLVLEKAVGLWIKIPCTDY
            IGSTCFHFCDVLDMLPTGPCEPLRTYGLPCHCPKFGTSLPKSEFVVPDLELP
            SWLTGNVRIESLVSSSGKRLGCIKIASLKG1"
ORIGIN
Alignment Scores:
Pred. No.: 5,44e-06 Length: 1043
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
Gaps: 0
DB: 9

US-10-030-937-68 (1-16) x HUMGM2A (1-1043)

Qy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
Db 190 TTTTCTGGGATAACTGTGATGAAGGAGGACCTGCGGTGATCAGA 237

RESULT 7
LOCUS CQ728078 1045 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 14012 from Patent WO02069579.
ACCESSION CQ728078
VERSION CQ728078.1 GI:42295943
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
Patent: WO 02068579-A 14012 06-SEP-2002;
PE Corporation (NY) (US)
LOCATION/Qualifiers
    1..1045
        /organism="Homo sapiens"
JOURNAL
FEATURES
    source
        1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
        190 TTTTCTGGGATAACTGTGATGAAGGAGGACCTGCGGTGATCAGA 237
```

```
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 5,45e-06 Length: 1045
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
Gaps: 0
DB: 6

US-10-030-937-68 (1-16) x CQ728078 (1-1045)

Qy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
Db 191 TTTTCTGGGATAACTGTGATGAAGGAGGACCTGCGGTGATCAGA 238

RESULT 8
LOCUS HSGM2A2 1047 bp DNA linear PRI 08-JUL-1999
DEFINITION Homo sapiens GM2 activator protein (GM2A) gene, exon 2, complete
sequence.
ACCESSION AF124718
VERSION AF124718.1 GI:4587476
KEYWORDS
SEGMENT 2 of 3
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1047)
AUTHORS Chen, B., Rigat, B., Curry, C. and Mahuran, D.J.
TITLE Structure of the GM2A gene: identification of an exon 2 nonsense
mutation and a naturally occurring transcript with an in-frame
deletion of exon 2
Am. J. Hum. Genet. 65 (1), 77-87 (1999)
JOURNAL 99294584
MEDLINE 10364519
PUBMED 10364519
REFERENCE 2 (bases 1 to 1047)
AUTHORS Chen, B., Rigat, B., Curry, C. and Mahuran, D.J.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1999) Structural Biology and Biochemistry,
Hospital For Sick Children, 555 University Ave., Toronto, Ontario
M5G1X8, Canada
FEATURES
    Location/Qualifiers
        1..1047
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            <1..390
            /gene="GM2A"
            /number=1
        391..552
            /gene="GM2A"
            /number=2
        553..>1047
            /gene="GM2A"
            /number=2
    intron
    exon
    intron

ORIGIN
Alignment Scores:
Pred. No.: 5,46e-06 Length: 1047
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
Gaps: 0
DB: 9

US-10-030-937-68 (1-16) x HSGM2A2 (1-1047)

Qy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
```


/gene="GM2A"
 /codon_start=1
 /product="GM2 ganglioside activator, precursor"
 /protein_id="AAH09273.1"
 /db_xref="GI:14424507"
 /db_xref="LOCUSID:2760"
 /db_xref="MIM:272750"
 /translation="MQSLMQAPLLIALGILLAPAAQHLKPKPSQLSFSFWDNCEGKD
 PAVIRSLTLEPDPVIVPGNVTLSVVGSTVSPKVDLVLEKEVAGLWIKIPCTDY
 IGSCFHFCDVLDMLIPTGEPCEPLRTYGLPCHCPKEGTYSLPKSEFVVPDLELP
 SWLTGNVRIESVLSGKRLGCIKIAASLKGI"

ORIGIN

Alignment Scores:
 Pred. No.: 1.34e-05 Length: 2413
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservatives: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 9 Gaps: 0

US-10-030-937-68 (1-16) x BC009273 (1-2413)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 Db 129 TTTTCTGGGATACTGTGATGAAGGAAGGACCTCGCGGTATCAGA 176

RESULT 11

AX330938 2436 bp DNA linear PAT 09-JAN-2002
 LOCUS Sequence 1447 from Patent WO0194629.
 DEFINITION AX330938
 ACCESSION AX330938
 VERSION AX330938.1 GI:18121572
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Young, F.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
 Horrigan, S., Soppet, D.R. and Weaver, Z.
 Cancer gene determination and therapeutic screening using signature
 gene sets

JOURNAL

Patent: WO 0194629-A 1447 13-DEC-2001;
 Avalon Pharmaceuticals (US)

FEATURES

Location/Qualifiers
 1..2436
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
 Pred. No.: 1.35e-05 Length: 2436
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservatives: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 6 Gaps: 0

US-10-030-937-68 (1-16) x AX330938 (1-2436)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 Db 158 TTTTCTGGGATACTGTGATGAAGGAAGGACCTCGCGGTATCAGA 205

RESULT 12

HSGM2APT 2436 bp mRNA linear PRI 15-FEB-1995
 LOCUS HSGM2APT
 DEFINITION H.sapiens mRNA for GM2 activator protein.
 ACCESSION X62078
 VERSION X62078.1 GI:313158
 KEYWORDS G(M2) activator protein.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

FEATURES

source

Location/Qualifiers

1..2436

/organism="Homo sapiens"

/mol_type="mRNA"

/isolate="patient with juvenile form of Sandhoff disease"

/db_xref="taxon:9606"

/clone="pUC18"

/cell_type="fibroblast"

/clone_lib="cDNA"

59..640

/note="alternative"

/codon_start=1

/product="GM2 activator protein"

/protein_id="CAA43993.1"

/db_xref="GI:673415"

/db_xref="GOA:P17900"

/db_xref="Swiss-Prot:P17900"

/translation="MQSLMQAPLLIALGILLATPAQHLKPKPSQLSFSFWDNCEGKD
 PAVIRSLTLEPDPVIVPGNVTLSVVGSTVSPKVDLVLEKEVAGLWIKIPCTDY
 IGSCFHFCDVLDMLIPTGEPCEPLRTYGLPCHCPKEGTYSLPKSEFVVPDLELP
 SWLTGNVRIESVLSGKRLGCIKIAASLKGI"

71..640

/note="alternative"

/codon_start=1

/product="GM2 activator protein"

/protein_id="CAA43994.1"

/db_xref="GI:673416"

/translation="MQAPLLIALGILLATPAQHLKPKPSQLSFSFWDNCEGKDPAVI
 RSLTLEPDPVIVPGNVTLSVVGSTVSPKVDLVLEKEVAGLWIKIPCTDYIGSC
 TPEHFCVDVLDMLIPTGEPCEPLRTYGLPCHCPKEGTYSLPKSEFVVPDLELP
 TGNVRIESVLSGKRLGCIKIAASLKGI"

ORIGIN

Alignment Scores:
 Pred. No.: 1.35e-05 Length: 2436
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservatives: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 9 Gaps: 0

US-10-030-937-68 (1-16) x HSGM2APT (1-2436)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 Db 158 TTTTCTGGGATACTGTGATGAAGGAAGGACCTCGCGGTATCAGA 205

RESULT 13

AC011391 107320 bp DNA linear HTG 23-APR-2001
 LOCUS AC011391
 DEFINITION Homo sapiens chromosome 5 clone CTF-176L22, WORKING DRAFT SEQUENCE,
 10 ordered pieces.
 ACCESSION AC011391

```

VERSION AC011391.5 GI:9256281
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 107320)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 107320)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7710539.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 132074, H341
Center clone name: CIT978SKB_176L22
-----
Summary Statistics
Consensus quality: 98512 bases at least Q40
Consensus quality: 10463 bases at least Q30
Consensus quality: 105879 bases at least Q20
Estimated insert size: 120000; pulse field gel estimation
Estimated insert size: 106870; sum-of-contigs estimation
Quality coverage: 6.0 in Q20 bases; pulse field gel estimation
Quality coverage: 6.73 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 7321: contig of 7321 bp in length
* 7422 7421: gap of unknown length
* 16470 16470: contig of 9049 bp in length
* 16571 16570: gap of unknown length
* 38101 38101: contig of 21531 bp in length
* 38201 38201: gap of unknown length
* 54501 54501: contig of 16300 bp in length
* 54601 54601: gap of unknown length
* 58262 58262: contig of 3661 bp in length
* 58263 58262: gap of unknown length
* 62481 62481: contig of 4119 bp in length
* 62581 62581: gap of unknown length
* 77487 77487: contig of 14906 bp in length
* 77588 77587: gap of unknown length
* 83729 83728: contig of 6141 bp in length
* 83828 83828: gap of unknown length
* 98662 98662: contig of 14834 bp in length
* 98663 98662: gap of unknown length
* 98763 107320: contig of 8558 bp in length.
*
FEATURES
Location/Qualifiers
1..107320
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-176L22"
/clone_lib="Caltech human BAC library B"
ORIGIN
Alignment Scores:
Pred. No.: 0.000788 Length: 107320

```

```

Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 2 Gaps: 0
US-10-030-937-68 (1-16) x AC011391 (1-107320)
Qy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
|||||
Db 22453 TTTTCTGGGATACCTGTGATGAAGGAGGACCTGCGGTGATCAGA 22500
|||||
RESULT 14
AC011342/c
LOCUS AC011342 120584 bp DNA linear PRI 24-OCT-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-276H5, complete sequence.
ACCESSION AC011342
VERSION AC011342.5 GI:16356867
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 120584)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 120584)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 120584)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS DOE Joint Genome Institute
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Oct 24, 2001 this sequence version replaced gi:9256276.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.3.
FEATURES
Location/Qualifiers
1..120584
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-276H5"
ORIGIN
Alignment Scores:
Pred. No.: 0.000893 Length: 120584
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 9 Gaps: 0
US-10-030-937-68 (1-16) x AC011342 (1-120584)
Qy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
|||||
Db 18898 TTTTCTGGGATACCTGTGATGAAGGAGGACCTGCGGTGATCAGA 18851
|||||
RESULT 15
AC008385/c
LOCUS AC008385 151712 bp DNA linear PRI 04-JUN-2002
DEFINITION Homo sapiens chromosome 5 clone CTC-224D3, complete sequence.
ACCESSION AC008385
VERSION AC008385.7 GI:21321772

```

KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151712)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151712)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 151712)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 151712)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jun 4, 2002 this sequence version replaced gi:14550298.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sbgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated total Number of Errors is 0.3.
FEATURES
source
1..151712
/location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-224D3"
ORIGIN
Alignment Scores:
Pred. No.: 0.00114 Length: 151712
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 9 Gaps: 0
US-10-030-937-68 (1-16) x AC008385 (1-151712)
QY 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
Db 99603 TTTTCTGGGATAACTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 99556
Search completed: November 18, 2004, 22:21:54
Job time : 674.378 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 16:05:15 ; Search time 71.9644 Seconds
(without alignments)
1167.114 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 94

Sequence: 1 FSWDNCFEGKDAVIR 16

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10030937/runat_16112004_153013_2903/app_query.fasta_1.789
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937 -CCN_1_1240 @runat_16112004_153013_2903 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 23Sep04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	90.4	579	4	Aaf54730 Nucleotid
2	85	90.4	953	12	Adq17711 Human sof
3	85	90.4	1043	4	Aaf54705 Nucleotid
4	85	90.4	1043	4	Aaf54708 Nucleotid
5	85	90.4	1047	4	Aaf54700 Nucleotid
6	85	90.4	1047	4	Aaf54703 Nucleotid

7	85	90.4	1047	4	Aaf54706	Aaf54706 Nucleotid
8	85	90.4	1935	10	ADB47402	ADB47402 Human cDN
9	85	90.4	2384	6	ABK34915	ABK34915 Human cDN
10	85	90.4	2436	3	AAC55714	AAC55714 Human GM2
11	85	90.4	2436	6	ABL63110	ABL63110 Breast ca
12	85	90.4	2436	10	ADD71046	ADD71046 Human GM2
13	85	90.4	2436	11	ADN95859	ADN95859 Human BEC
14	85	90.4	2471	5	AAS64907	AAS64907 DNA encod
15	85	90.4	2478	6	ABV78068	ABV78068 Hypoxia-r
16	85	90.4	2478	12	ADN03619	ADN03619 Antipsori
17	85	90.4	2498	5	AAS81113	AAS81113 DNA encod
18	85	90.4	3988	12	ADQ22367	ADQ22367 Human sof
19	82	87.2	279	3	AAA41272	AAA41272 Human sec
20	82	87.2	368	3	AAA42669	AAA42669 Human sec
21	82	87.2	1983	2	AAT61025	AAT61025 Rat GM2 a
22	82	87.2	1983	10	ADB52361	ADB52361 Primary r
23	76	80.9	579	4	Aaf54698	Aaf54698 Nucleotid
24	58	61.7	139	12	ADOL1344	ADOL1344 SNP targe
25	57	60.6	380	4	AAL02289	AAL02289 Human rep
26	57	60.6	577	6	ABQ54556	ABQ54556 Human ova
27	51	54.3	551	11	ACH96897	ACH96897 Klebsiell
28	50	53.2	1178	4	AAL61284	AAL61284 Human pol
29	50	53.2	2014	4	AAL59498	AAL59498 Human pol
30	50	53.2	2243	11	ADM01616	ADM01616 Human cDN
31	50	53.2	2367	4	ABL22285	ABL22285 Drosophil
32	50	53.2	2367	12	ADO07755	ADO07755 Fly polyn
33	50	53.2	2452	10	ADA52519	ADA52519 Human cod
34	50	53.2	2459	8	ABV75073	ABV75073 Drosophil
35	50	53.2	10386	4	ABL22264	ABL22264 Drosophil
36	49	52.1	6306	6	AA040757	AA040757 Human kin
37	49	52.1	6829	10	ACF80031	ACF80031 Modifier
38	49	52.1	6629	10	ADL15089	ADL15089 Human mal
39	49	52.1	6629	10	ADL15089	ADL15089 Human mal
40	49	52.1	6899	4	AAK52092	AAK52092 Human pol
41	49	52.1	7018	4	AAK53076	AAK53076 Human pol
42	49	52.1	7664	10	ADF14281	ADF14281 Human end
43	49	52.1	7676	6	ABN85383	ABN85383 Human NOV
44	49	52.1	7728	10	ADF14280	ADF14280 Human end
45	49	52.1	7785	4	AAS06703	AAS06703 Polynucle

ALIGNMENTS

RESULT 1
AAF54730
ID AAF54730 standard; DNA; 579 BP.

AC AAF54730;

XX 15-MAY-2001 (first entry)

DT 15-MAY-2001 (first entry)

XX Nucleotide sequence of a human polynucleotide sequence.

DE Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

XX Homo sapiens.

OS Homo sapiens.

XX WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INNR) BIOMERIEUX STELHYS.

PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX

DR WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and

PT autoimmune diseases, particularly multiple sclerosis, using specified

PT polypeptides or related nucleic acid or ligand.

XX Claim 11; Page 208; 209pp; French.

XX The present sequence represents a human polynucleotide sequence, which is

CC used in the method of the invention. The specification describes a method

CC which uses at least one polypeptide or polynucleotide sequence belonging

CC to the perlecan, precursor of the retinol-binding plasma protein,

CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B

CC protein families. The method is used for detecting, preventing or

CC treating a degenerative, neurological and/or auto-immune disease. The

CC polynucleotides and polypeptides are used for diagnosis, prognosis,

CC prevention and treatment of multiple sclerosis (in its various forms and

CC phases). They may also be useful in cases of e.g. Alzheimer's and

CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid

CC polyarthritis and lupus erythematosus, including use as vaccines and in

CC gene therapy (expression of sense or antisense sequences). They can also

CC be used to assess efficacy of potential therapeutic agents, particularly

CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 579 BP; 83 A; 66 C; 89 G; 82 T; 0 U; 259 Other;

SQ

Alignment Scores: 1.87e-05 Length: 579

Pred. No.: 85.00 Matches: 14

Score: 93.33% Conservative: 0

Percent Similarity: 93.33% Mismatches: 1

Best Local Similarity: 90.43% Indels: 0

Query Match: 4 Gaps: 0

DB:

US-10-030-937-68 (1-16) x AAF54730 (1-579)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIle 15

DB 100 TTYNSNTGGGAYAAVTGTYTGARGNAARGAAYCCNGCNGTNATH 144

RESULT 2

ADQ17711

ID ADQ17711 standard; DNA; 953 BP.

XX AC ADQ17711;

XX DT 26-AUG-2004 (first entry)

XX DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 528.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;

KW ds.

XX Homo sapiens.

OS

XX WO2004048938-A2.

PN

XX 10-JUN-2004.

PD

XX 26-NOV-2003; 2003WO-US038193.

PF

XX 26-NOV-2002; 2002US-0429739P.

PR

XX (PROT-) PROTEIN DESIGN LABS INC.

PA

XX Aziz N, Ginsburg WM, Zlotnik A;

PI

XX WPI; 2004-441208/41.

DR

XX Early detection of soft tissue sarcoma comprises determining expression

PT of a gene in a first soft tissue sample and a normal soft tissue sample

PT and comparing the gene expression, also useful in treating soft tissue

PT sarcoma.

XX Example 2; SEQ ID NO 528; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma

CC which comprises obtaining a first soft tissue sample from an individual

CC and a normal soft tissue sample from the same or different individual,

CC determining the expression of a gene in both samples and comparing the

CC expression of the gene in both soft tissue samples, where a higher level

CC of protein expression in the first soft tissue sample indicates the

CC presence of soft tissue sarcoma. The method of the invention has

CC cytostatic applications and may be useful for detecting soft tissue

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic

CC acid sequences may be useful in diagnostic and screening applications.

CC The current sequence is that of a human soft tissue sarcoma-upregulated

CC DNA of the invention. The current sequence is not shown within the

CC specification per se but was submitted in CD format by the inventor.

XX Sequence 953 BP; 201 A; 286 C; 231 G; 235 T; 0 U; 0 Other;

SQ

Alignment Scores: 3.34e-05 Length: 953

Pred. No.: 85.00 Matches: 15

Score: 93.75% Conservative: 0

Percent Similarity: 93.75% Mismatches: 1

Best Local Similarity: 90.43% Indels: 0

Query Match: 12 Gaps: 0

DB:

US-10-030-937-68 (1-16) x ADQ17711 (1-953)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16

DB 190 TTTTCCTGGGATACCTGTGTGATGAAGGAAGGACCCCTGGGTGATCAGA 237

RESULT 3

AAF54705

ID AAF54705 standard; DNA; 1043 BP.

XX AC AAF54705;

XX DT 15-MAY-2001 (first entry)

XX DE Nucleotide sequence of a human polynucleotide sequence.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;

KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;

KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;

KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;

KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

XX Homo sapiens.

OS

XX WO200105422-A2.

PN

XX 25-JAN-2001.

PD

XX 17-JUL-2000; 2000WO-FR002057.

PF

XX 15-JUL-1999; 99FR-00009372.

PR

XX (INMR) BIOMERIEUX STELHYS.

PA

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

PI

XX WPI; 2001-159475/16.

DR

XX Detecting, preventing and treating degenerative, neurological and

PT autoimmune diseases, particularly multiple sclerosis, using specified

PT polypeptides or related nucleic acid or ligand.

XX Claim 11; Page 181-182; 209pp; French.

PS The present sequence represents a human polynucleotide sequence, which is

XX used in the method of the invention. The specification describes a method

CC

CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 CC
 SQ Sequence 1043 BP; 222 A; 303 C; 247 G; 271 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3.71e-05 Length: 1043
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservativity: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-68 (1-16) x AAF54705 (1-1043)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 Db 190 TTTTCTGGGATAACTGTGATGAAGGGAAGACCCCTGCGGTGATCAGA 237

RESULT 4

AAF54708

ID AAF54708 standard; DNA; 1043 BP.

XX AC AAF54708;

XX DT 15-MAY-2001 (first entry)

XX DE Nucleotide sequence of a human polynucleotide sequence.

XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99PR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX DR WPI; 2001-159475/16.

XX PT Detecting, preventing and treating degenerative, neurological and
 XX autoimmune diseases, particularly multiple sclerosis, using specified
 XX polypeptides or related nucleic acid or ligand.

XX PS Claim 11; Page 183; 209pp; French.

XX CC The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B

CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 CC
 SQ Sequence 1043 BP; 222 A; 303 C; 247 G; 271 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.71e-05 Length: 1043
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservativity: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-68 (1-16) x AAF54708 (1-1043)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16

Db 190 TTTTCTGGGATAACTGTGATGAAGGGAAGACCCCTGCGGTGATCAGA 237

RESULT 5

AAF54700

ID AAF54700 standard; DNA; 1047 BP.

XX AC AAF54700;

XX DT 15-MAY-2001 (first entry)

XX DE Nucleotide sequence of a human polynucleotide sequence.

XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99PR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX DR WPI; 2001-159475/16.

XX PT Detecting, preventing and treating degenerative, neurological and
 XX autoimmune diseases, particularly multiple sclerosis, using specified
 XX polypeptides or related nucleic acid or ligand.

XX PS Claim 11; Page 179; 209pp; French.

XX CC The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,

CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX SQ Sequence 1047 BP; 244 A; 227 C; 286 G; 290 T; 0 U; 0 Other;

Alignment Scores: 3.73e-05 Length: 1047
 Pred. No.: 85.00 Matches: 15
 Score: 93.75% Conservativity: 0
 Percent Similarity: 93.75% Mismatches: 1
 Best Local Similarity: 90.43% Indels: 0
 Query Match: 4 Gaps: 0
 DB:

US-10-030-937-68 (1-16) x AAF54700 (1-1047)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 |||||
 Db 409 TTTTCTGGGATAACTGTGATGAAGGGAAGGACCTCGCGTGATCAGA 456

RESULT 6

AAF54703
 ID AAF54703 standard; DNA; 1047 BP.

XX AC AAF54703;

XX DT 15-MAY-2001 (first entry)

XX DE Nucleotide sequence of a human polynucleotide sequence.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99FR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX DR WPI; 2001-159475/16.

XX PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX PS Claim 11; Page 180; 209pp; French.

XX The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid

CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX SQ Sequence 1047 BP; 244 A; 227 C; 286 G; 290 T; 0 U; 0 Other;

Alignment Scores: 3.73e-05 Length: 1047
 Pred. No.: 85.00 Matches: 15
 Score: 93.75% Conservativity: 0
 Percent Similarity: 93.75% Mismatches: 1
 Best Local Similarity: 90.43% Indels: 0
 Query Match: 4 Gaps: 0
 DB:

US-10-030-937-68 (1-16) x AAF54703 (1-1047)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 |||||
 Db 409 TTTTCTGGGATAACTGTGATGAAGGGAAGGACCTCGCGTGATCAGA 456

RESULT 7

AAF54706

ID AAF54706 standard; DNA; 1047 BP.

XX AC AAF54706;

XX DT 15-MAY-2001 (first entry)

XX DE Nucleotide sequence of a human polynucleotide sequence.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99FR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX DR WPI; 2001-159475/16.

XX PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX PS Claim 11; Page 182; 209pp; French.

XX The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly

CC compounds that reduce or inhibit toxicity towards glial cells
 XX SQ Sequence 1047 BP; 244 A; 227 C; 286 G; 290 T; 0 U; 0 Other;

Alignment Scores: 3.73e-05 Length: 1047
 Pred. No.: 85.00 Matches: 15
 Score: 85.00
 Percent Similarity: 93.75%
 Best Local Similarity: 93.75%
 Query Match: 90.43%
 Indels: 0
 Gaps: 0

US-10-030-937-68 (1-16) x AAF54706 (1-1047)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 DB 409 TTTTCTGGGATAACTGTGTATGAAGGAGGACCTCGGTGATCAGA 456

RESULT 8
 ADB47402
 ID ADB47402 standard; cDNA; 1935 BP.
 AC ADB47402;
 XX 04-DEC-2003 (first entry)
 DT
 DE Human cDNA upregulated in dendritic cells SEQ ID NO 102.

XX ss; gene; human; dendritic cells; high throughput; cancer;
 KW infectious disease; autoimmune disease; allergy;
 KW graft versus host disease; vaccine enhancing; gene therapy.
 XX Homo sapiens.

XX US2003134283-A1.
 XX 17-JUL-2003.

XX 03-OCT-2001; 2001US-00971392.
 XX 03-OCT-2000; 2000US-0237652P.

XX (PETE/) PETERSON D P.
 XX (PEAR/) PEARSON C I.
 XX (COCK/) COCKS B G.

XX Peterson DP, Pearson CI, Cocks BG;
 XX WPI; 2003-662509/62.

XX New combination comprises cDNAs that are differentially expressed in
 PT dendritic cells useful for preparing a composition for diagnosing or
 PT treating cancer, infectious disease, autoimmunity, allergy or graft
 PT versus host disease.

XX Claim 1; SEQ ID NO 102; 28pp; English.

XX The invention relates to a combination comprising cDNAs that are
 CC differentially expressed in dendritic cells (DC). Also included is a high
 CC throughput method for detecting differential expression of one or more
 CC cDNAs in a sample containing nucleic acids. The combination is useful for
 CC preparing a composition for diagnosing, treating and monitoring the
 CC treatment of cancer, infectious disease, autoimmunity, allergy or graft
 CC versus host disease, or for enhancing a vaccine. The present sequence
 CC represents a human cDNA upregulated in dendritic cells. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?DocID=20030134283.

XX SQ Sequence 1935 BP; 459 A; 507 C; 460 G; 509 T; 0 U; 0 Other;
 Alignment Scores: 7.63e-05 Length: 1935

Score: 85.00 Matches: 15
 Percent Similarity: 93.75%
 Best Local Similarity: 93.75%
 Query Match: 90.43%
 Indels: 0
 Gaps: 0

US-10-030-937-68 (1-16) x ADB47402 (1-1935)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 DB 201 TTTTCTGGGATAACTGTGTATGAAGGAGGACCTCGGTGATCAGA 248

RESULT 9
 ABK34915
 ID ABK34915 standard; cDNA; 2384 BP.
 XX
 AC ABK34915;
 XX
 DT 08-MAY-2002 (first entry)

XX Human cDNA encoding secreted protein #53.
 XX Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
 KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
 KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
 KW tissue regeneration; wound healing; burn; haematopoiesis;
 KW myeloid cell deficiency; lymphoid cell deficiency.

XX Homo sapiens.
 XX WO200177288-A2.
 XX 18-OCT-2001.

XX 29-MAR-2001; 2001WO-US010224.
 XX 06-APR-2000; 2000US-0195582P.

XX (GEMY) GENETICS INST INC.
 XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
 XX Gulukota K, Graham JR;

XX WPI; 2002-179321/23.

XX Five hundred and ninety two polynucleotides derived from a variety of
 PT human tissue sources which encode secreted proteins, useful for treating
 PT immune deficiencies and disorders such as autoimmune disorders.

XX Claim 1; Page 95-96; 372pp; English.

XX The invention relates to 592 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins. The polynucleotides can be used as probes for the
 CC identification and isolation of full length cDNA and genomic DNA. The
 CC polynucleotides and proteins can also be used as nutritional supplements.
 CC The proteins are useful in the treatment of various immune deficiencies
 CC and disorders such as viral infections, bacterial infections, fungal
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
 CC and conditions (e.g. asthma). They are also useful for treating
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
 CC useful for tissue regeneration, for wound healing and in the treatment of
 CC burns, incisions and ulcers. The proteins are also useful for regulating
 CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
 CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention
 XX SQ Sequence 2384 BP; 550 A; 618 C; 571 G; 645 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 9,74e-05 Length: 2384
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 6 Gaps: 0

US-10-030-937-68 (1-16) x ABK34915 (1-2384)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 |||||
 Db 112 TTTTCTGGGATAACTGTGATGAAGGAAGGACCCCTGGGTGATCAGA 159
 |||||

RESULT 10
 AAC55714
 ID AAC55714 standard; cDNA; 2436 BP.
 XX AC AAC55714;
 XX AC AAC55714;
 DT 17-JAN-2001 (first entry)
 DE Human GM2 activator protein cDNA sequence from Genbank X62078.
 XX Human; differentially regulated gene; macrophage development; diagnosis;
 KW matrix metalloproteinase 19; MMP19; antiarthritic; antiinflammatory;
 KW destructive macrophage development inhibitor; arthritis;
 KW colorectal cancer; immune response; ss.
 XX Homo sapiens.
 OS
 XX WO2000055373-A2.
 PN
 XX 21-SEP-2000.
 PD
 XX 15-MAR-2000; 2000WO-US006883.
 PF
 XX 15-MAR-1999; 99US-0124530P.
 PR
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 FA
 XX Murray R;
 FI
 XX WPI; 2000-628200/60.
 DE
 XX Screening drug candidates comprises adding a drug to a cell expressing an
 PT expression profile gene and determining the effect of the drug on the
 PT expression of the expression profile gene.
 XX
 XX Claim 1; Page; 99pp; English.
 CC The present invention describes a method for screening drug candidates.
 CC The method comprises adding a drug to a cell that expresses an expression
 CC profile gene encoding a protein encoded by 5 sequences of defined base
 CC pairs as given in C55638, C55642, C55643, C55644 and C55653 or a sequence
 CC represented by Genbank accession number X92521, X62466, J04130, X62087
 CC and X76534 (or a fragment) and determining the effect of the drug on the
 CC expression of the expression profile gene. An inhibitor of matrix
 CC metalloproteinase 19 (MMP-19), preferably an antibody, is useful for
 CC treating destructive macrophage disorders (DMD) by inhibiting DM
 CC development in a cell of an individual having arthritis. Antibodies to
 CC MMP-19 are useful for localising a therapeutic moiety preferably
 CC cytotoxic agent or a radioisotope to colorectal cancer tissue. A
 CC composition comprising MMP-19 is useful for eliciting an immune response
 CC in an individual. C55635 to C55710 represent human differentially
 CC regulated genes of the invention. The present sequence represents the
 CC human GM2 activator protein cDNA sequence according to the Genbank
 CC accession number X62078. N.B. The present sequence is not given in the
 CC present specification, but it is specifically claimed by its Genbank
 CC accession number
 XX Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 9,99e-05 Length: 2436
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 3 Gaps: 0

US-10-030-937-68 (1-16) x AAC55714 (1-2436)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 |||||
 Db 158 TTTTCTGGGATAACTGTGATGAAGGAAGGACCCCTGGGTGATCAGA 205
 |||||

RESULT 11
 ABL63110
 ID ABL63110 standard; DNA; 2436 BP.
 XX AC ABL63110;
 XX AC ABL63110;
 DT 15-MAY-2002 (first entry)
 DE Breast cancer related gene sequence SEQ ID NO:1447.
 XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cycostatic; gene therapy; antineoplastic; wilm's tumour; adenocarcinoma;
 KW gene; ds.
 XX Homo sapiens.
 OS
 XX WO200194629-A2.
 PN
 XX 13-DEC-2001.
 PD
 XX 30-MAY-2001; 2001WO-US010838.
 PF
 XX 05-JUN-2000; 2000US-0209473P.
 PR 05-JUN-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-0233133P.
 PR 18-SEP-2000; 2000US-0233617P.
 PR 20-SEP-2000; 2000US-0234009P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 20-SEP-2000; 2000US-0234052P.
 PR 22-SEP-2000; 2000US-0234509P.
 PR 22-SEP-2000; 2000US-0234567P.
 PR 25-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.

```
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
XX Claim 1; SEQ ID NO 1447; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
XX Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9.99e-05 Length: 2436
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-68 (1-16) x ABL63110 (1-2436)
QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 158 TTTTCTGGGATAACTGTGTGATGAAGGAGGACCTCGGTGATCAGA 205

RESULT 12
ADD71046
ID ADD71046 standard; DNA; 2436 BP.
XX
XX ADD71046;
XX
XX 15-JAN-2004 (first entry)
XX
XX Human GM2 ganglioside activated protein gene SEQ ID NO:50.
XX
XX liver cancer; chronic hepatitis; cirrhosis; liver disease; hepatotropic;
XX cytostatic; gene therapy; human; gene; ds.
XX
XX Homo sapiens.
XX
XX W02003061564-A2.
XX
XX 31-JUL-2003.
XX

20-DEC-2002; 2002WO-US040718.
XX
XX 21-DEC-2001; 2001US-0341815P.
PR
PR 31-DEC-2001; 2001US-0343185P.
XX
XX (GENE-) GENE LOGIC INC.
PA (LGBI-) LG BIOMEDICAL INST.
XX
XX Koh SS, Liu Q, Chung H, Zeng W, Lee B, Yerawilli S, Song SY;
PI WPI; 2003-663343/62.
XX
XX Diagnosing liver cancer cells, useful for treating liver cancer
PT associated with chronic hepatitis or cirrhosis comprises detecting the
PT level of expression in a tissue sample of one or more genes associated
PT with cancerous liver tissues.
XX
XX Claim 1; SEQ ID NO 50; 176pp; English.
XX
XX The present invention describes a method for diagnosing liver cancer
CC cells comprising detecting the level of expression in a tissue sample of
CC one or more genes given in the specification (see ADD70997 to ADD71105),
CC where differential expression of the genes is indicative of liver cancer.
CC Also described: (1) detecting the progression of liver cancer in a
CC patient; (2) monitoring the treatment of a patient with liver cancer; (3)
CC treating a patient with liver cancer; (4) typing a liver disease in a
CC patient; (5) detecting the presence or progression of liver cancer in a
CC patient with chronic hepatitis or cirrhosis; (6) differentiating liver
CC cancer related to chronic hepatitis from liver cancer related to
CC cirrhosis; (7) screening for an agent capable of modulating the onset or
CC progression of liver cancer; (8) a composition comprising at least two
CC oligonucleotides comprising a sequence that specifically hybridises to
CC any of the genes; (9) a solid support comprising the at least two
CC oligonucleotides; (10) a computer system comprising a database containing
CC information identifying the level in liver tissue of a set of genes; (11)
CC a method for using the computer system to present information identifying
CC the expression level in tissue or cell of any of the genes; and (12) a
CC therapeutic agent for slowing or halting the progression of liver cancer.
CC The methods are useful for treating liver cancer associated with chronic
CC hepatitis or cirrhosis. The present sequence represents a specifically
CC claimed human gene sequence which is used in the exemplification of the
CC present invention.
XX
XX Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9.99e-05 Length: 2436
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 10 Gaps: 0

US-10-030-937-68 (1-16) x ADD71046 (1-2436)
QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 158 TTTTCTGGGATAACTGTGTGATGAAGGAGGACCTCGGTGATCAGA 205

RESULT 13
ADN95859
ID ADN95859 standard; DNA; 2436 BP.
XX
XX ADN95859;
XX
XX 01-JUL-2004 (first entry)
XX
XX Human BEC/LEC-related gene sequence SeqID783.
XX
XX growth; differentiation; blood endothelial cell; BEC;
XX lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
XX lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;
XX
```

KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
 KW inflammatory disease; cancer metastasis; lymphatic system; gene; ds;
 XX human.
 XX Homo sapiens.
 XX WO2003080640-A1.
 XX 02-OCT-2003.
 XX 07-MAR-2003; 2003WO-US006900.
 XX 07-MAR-2002; 2002US-0363019P.
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX (LIGN) LICENTIA LTD.
 XX Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
 XX WPI; 2003-876999/81.
 XX P-PSDB; ADN95858.
 XX Example 1; SEQ ID NO 783; 176pp; English.
 XX This invention relates to a method of differentially modulating the
 CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
 CC endothelial cells (LEC) comprises contacting endothelial cells with a
 CC composition comprising an agent that differentially modulates blood or
 CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
 CC identifying a human subject with lymphoedema and with a mutation in at
 CC least one allele of a gene encoding a LEC protein, where the mutation
 CC correlates with lymphoedema in human subjects, and with the proviso that
 CC the LEC protein is not VEGFR-3; and administering to the subject a
 CC composition comprising a lymphatic growth agent selected from VEGF-C or
 CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
 CC the development of compounds with an antiangiogenic, cytostatic,
 CC vasotropic or antiinflammatory activity or for gene therapy. The method
 CC is useful in modulating the growth or differentiation of blood
 CC endothelial cells or lymphatic endothelial cells, in treating hereditary
 CC lymphoedema, in screening for an endothelial cell disorder or
 CC predisposition to the disorder or in monitoring the efficacy or toxicity
 CC of a drug on endothelial cells. The agent is useful in manufacturing a
 CC medicament for the differential modulation of blood vessel endothelial
 CC cell or lymphatic vessel endothelial cell growth or differentiation. The
 CC lymphatic growth agent may also be used in manufacturing a medicament for
 CC the treatment of hereditary lymphoedema resulting from a mutation in a
 CC LEC gene or of other diseases involving the lymphatic vessels, such as
 CC various inflammatory diseases and cancer metastasis via the lymphatic
 CC system. The present sequence is that of a human LEC/BEC differentially
 CC expressed gene which is related to the method of the invention. Note: This
 CC sequence does not appear in the specification but was obtained by the
 XX indexer using the source data given in table 14 of the specification.
 XX SQ Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 9.99e-05 Length: 2436
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 11 Gaps: 0
 US-10-030-937-68 (1-16) x ADN95859 (1-2436)
 Qy 1 PheSerTrpAspAenCysPheGluGlyLysAspProAlaValIleArg 16
 Db 158 TTTTCTGGGATAACTGTGATGAAGGAAGGACCTCGCGGTGATCAGA 205
 RESULT 14
 ID AAS64907
 AD AAS64907 standard; cDNA; 2471 BP.
 XX

AC AAS64907;
 XX 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #711.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX P-PSDB; ABG00720.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 711; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 2471 BP; 578 A; 639 C; 583 G; 671 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.000102 Length: 2471
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 5 Gaps: 0
 US-10-030-937-68 (1-16) x AAS64907 (1-2471)
 Qy 1 PheSerTrpAspAenCysPheGluGlyLysAspProAlaValIleArg 16
 Db 192 TTTTCTGGGATAACTGTGATGAAGGAAGGACCTCGCGGTGATCAGA 239
 RESULT 15

ABV78068
ID ABV78068 standard; DNA; 2478 BP.
AC
XX ABV78068;
XX
XX 12-NOV-2002 (first entry)
XX
XX Hypoxia-regulated protein coding sequence #88.
XX
XX Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;
KW antiinflammatory; vulnery; gynecological; ophthalmological; vaccine;
KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;
KW preeclampsia; atherosclerosis; inflammatory condition; wound healing;
XX inflammation; erythropoiesis; hair loss; human; gene; ds.
XX
OS Homo sapiens.
XX
XX WO200246465-A2.
PN
XX
XX 13-JUN-2002.
XX
XX 10-DEC-2001; 2001WO-GB005458.
XX
XX 08-DEC-2000; 2000GB-00030076.
PR
XX 08-FEB-2001; 2001GB-00003156.
PR
XX 25-OCT-2001; 2001GB-00025666.
XX
XX (OXFO-) OXFORD BIOMEDICA UK LTD.
PA
XX
XX White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;
PI Rayner WN;
PI
XX
XX WPI; 2002-627238/67.
XX
XX Identifying a gene involved in disease for treating hypoxia-regulated
PT conditions, comprises comparing the transcriptome/proteome of two cell
PT types under different conditions and identifying a differentially
PT regulated gene.
XX
XX Claim 37; Page 397-398; 538pp; English.
PS
XX
XX The present invention relates to methods for identifying genes and
CC proteins that are implicated in a specific disease or physiological
CC condition. The method comprises comparing the transcriptome/proteome of a
CC specialised cell type implicated in a disease or condition with that of a
CC second specialised cell type, under two experimental conditions, and
CC identifying a gene that is differentially regulated in the two
CC specialised cell types under experimental conditions. ABV7873-ABV78116
CC and ABP65061-ABP65257 were identified using the methods of the invention.
CC The coding sequences and proteins are useful for treating a disease in a
CC patient, for manufacture of a medicament for treating hypoxia-regulated
CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
CC biological response to hypoxia conditions, or hypoxic-associated
CC pathology in a patient. The coding sequences and proteins are also useful
CC for monitoring the therapeutic treatment of a disease or physiological
CC condition, such as cancer, ischaemic conditions, reperfusion injury,
CC retinopathy, neonatal stress, preeclampsia, atherosclerosis, inflammatory
CC conditions, wound healing, inflammation, erythropoiesis or hair loss
XX
SQ Sequence 2478 BP; 588 A; 538 C; 586 G; 666 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.000102 Length: 2478
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 6 Gaps: 0
US-10-030-937-68 (1-16) x ABV78068 (1-2478)
QY 1 PheSerTrpAspAsnCysPheGluGlyAspProLalValIleArg 16

Db 195 TTTTCTGGGATAACTGTGATGAAGGAGACCCCTGCGTGATCAGA 242
Search completed: November 18, 2004, 19:48:42
Job time : 74.9644 secs

This Page Blank (uspto)


```
; Sequence 1, Application US/07841654B
; Patent No. 5260209
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Ibraghimov-Beskovnaya, Oxana
; APPLICANT: Ervasti, James M.
; APPLICANT: Leveille, Cynthia J.
; APPLICANT: Matsumura, Kiichiro
; TITLE OF INVENTION: DNA ENCODING DYSTROPHIN-ASSOCIATED
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/841.654B
; FILING DATE: 19920220
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: UIRF89-11AA
; TELEPHONE: 617 861-9540
; TELEFAX: 617 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..2855
; US-07-841-654B-1

Alignment Scores:
Pred. No.: 87.4 Length: 4200
Score: 48.00 Matches: 7
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 58.33% Mismatches: 3
Query Match: 51.06% Indels: 0
DB: 1 Gaps: 0

US-10-030-937-68 (1-16) x US-07-841-654B-1 (1-4200)
Qy 3 TripAspAsnCysPheGluGlyLysAspProAlaVal 14
Db 73 TGGAGCAGGTGTGCAGAGGGTGGAGACCCGGCTCTG 108

RESULT 3
US-07-946-234A-1
; Sequence 1, Application US/07946234A
; Patent No. 5308752
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Matsumura, Kiichiro
; TITLE OF INVENTION: DIAGNOSIS OF AUTOSOMAL MUSCULAR DYSTROPHY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS
; STREET: 2 Militia Drive
; CITY: Lexington
```

```
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/946,234A
; FILING DATE: 19920914
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: UIRF89-11AAA
; TELEPHONE: (617)861-6240
; TELEFAX: (617)861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..2855
; US-07-946-234A-1

Alignment Scores:
Pred. No.: 87.4 Length: 4200
Score: 48.00 Matches: 7
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 58.33% Mismatches: 3
Query Match: 51.06% Indels: 0
DB: 1 Gaps: 0

US-10-030-937-68 (1-16) x US-07-946-234A-1 (1-4200)
Qy 3 TripAspAsnCysPheGluGlyLysAspProAlaVal 14
Db 73 TGGAGCAGGTGTGCAGAGGGTGGAGACCCGGCTCTG 108

RESULT 4
US-08-123-161A-1
; Sequence 1, Application US/08123161A
; Patent No. 5449616
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Roberts, Steven L.
; APPLICANT: Anderson, Richard D.
; APPLICANT: Ibraghimov, Oxana B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,161A
; FILING DATE: 16-SEP-93
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/946,234
; FILING DATE: 14-SEP-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIRf89-11A4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..2855
; US-08-123-161A-1
;
; Alignment Scores:
; Pred. No.: 87.4 Length: 4200
; Score: 48.00 Matches: 7
; Percent Similarity: 75.00% Conservative: 2
; Best Local Similarity: 58.33% Mismatches: 3
; Query Match: 51.06% Indels: 0
; DB: 1 Gaps: 0
;
; US-10-030-937-68 (1-16) x US-08-123-161A-1 (1-4200)
;
; QY 3 TtpAspAsnCysPheGluGlyLysAspProAlaVal 14
; DB 73 TGGAGCAGGTGTGCAGAGGTGAGGACCGGCTCTG 108
;
; RESULT 5
; US-08-483-278-1
; Sequence 1, Application US/08483278
; Patent No. 5686073
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Ibraghimov, Oxana B.
; APPLICANT: Ervasti, James M.
; APPLICANT: Leveille, Cynthia J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,278
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/946,234
; FILING DATE: 14-SEP-92
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIRf89-11A5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..2855
; PCT-US93-01560-1
;
; US-10-030-937-68 (1-16) x US-08-483-278-1 (1-4200)
;
; QY 3 TtpAspAsnCysPheGluGlyLysAspProAlaVal 14
; DB 73 TGGAGCAGGTGTGCAGAGGTGAGGACCGGCTCTG 108
;
; RESULT 6
; PCT-US93-01560-1
; Sequence 1, Application PC/TUS9301560
; GENERAL INFORMATION:
; APPLICANT: University of Iowa Research Foundation
; TITLE OF INVENTION: DNA ENCODING DYSTROPHIN-ASSOCIATED
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01560
; FILING DATE: 19930219
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,654
; FILING DATE: 20-FEB-1992
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: UIRf89-11AA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 861-6240
; TELEFAX: 617 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..2855
; PCT-US93-01560-1
```



```
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 573
; LENGTH: 3940
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Inocyte ID No. 6673549 296811.10
US-09-976-594-573

Alignment Scores:
Pred. No.: 288 Length: 3940
Score: 45.00 Matches: 9
Percent Similarity: 52.17% Conservative: 3
Best Local Similarity: 39.13% Mismatches: 1
Query Match: 47.87% Indels: 10
DB: 4 Gaps: 1

US-10-030-937-68 (1-16) x US-09-976-594-573 (1-3940)

QY 3 TrpAspAsnCysPhe-----GlucLyLysaspPro 12
Db 354 TGGGATCACTGCTGGGGCCACCGGGCCAAAGCTAGCTCGGATGAGAGGAGTTGATCCT 413
QY 13 AlaValIle 15
Db 414 GCTGTTCTG 422

RESULT 11
US-09-341-587-7
; Sequence 7, Application US/09341587
; Patent No. 6346606
; GENERAL INFORMATION:
; APPLICANT: Mollenhauer, Jan
; TITLE OF INVENTION: Protein Containing an SRCR Domain
; FILE REFERENCE: 4121-108
; CURRENT APPLICATION NUMBER: US/09/341,587
; CURRENT FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: PCT/DE98/00096
; EARLIER FILING DATE: 1998-01-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 28720
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-341-587-7

Alignment Scores:
Pred. No.: 3-23e+03 Length: 28720
Score: 45.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 47.87% Indels: 0
DB: 3 Gaps: 0

US-10-030-937-68 (1-16) x US-09-341-587-7 (1-28720)

QY 1 PheSerTrpAspAsnCysPheGluGlyLys 10
Db 11659 TACAGTTGTGACAACTGTTTCCCAAGGAAAA 11688

RESULT 12
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
```

```
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (225001)..(240000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (240001)..(255000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (255001)..(270000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (270001)..(285000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (285001)..(300000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (300001)..(315000)
```

```

; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (315001)..(330000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (330001)..(345000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (450001)..(465000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (465001)..(480000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (480001)..(495000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (495001)..(510000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (510001)..(525000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (525001)..(540000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (540001)..(555000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (555001)..(570000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (570001)..(585000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t

```

```

; NAME/KEY: misc feature
; LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (690001)..(705000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (705001)..(720000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (720001)..(735000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (735001)..(750000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (750001)..(765000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (765001)..(780000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (780001)..(795000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (795001)..(810000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (810001)..(825000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (825001)..(840000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature

```

```

Alignment Scores:
Pred. No.: 2.85e+05 Length: 1230025
Score: 45.00 Matches: 6
Percent Similarity: 71.43% Conservative: 4
Best Local Similarity: 42.86% Mismatches: 4
Query Match: 47.87% Indels: 0
DB: Gaps: 0

```

US-10-030-937-68 (1-16) x US-09-198-452A-1 (1-1230025)

```

Oy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaVal 14
    :|||||:|||||:|||||:|||||:|||||:|||||:
Db 1035081 TATAGCTGGCGGCTGCTTCAATAAACTGATCCAGTATT 1035122

RESULT 13
US-10-140-002-65
; Sequence 65, Application US/10140002
; Patent No. 6725730
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc

```

```

; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C59
; CURRENT APPLICATION NUMBER: US/10/140,002
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 65
; LENGTH: 1606
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-002-65

Alignment Scores:
Pred. No.: 148 Length: 1606
Score: 44.00 Matches: 7
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 46.67% Mismatches: 5
Query Match: 46.81% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-68 (1-16) x US-10-140-002-65 (1-1606)
QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValle 15
Db 480 TTCTCTTGGACACACATCACTGATTCATTGATCTGTCACCCCTG 524

RESULT 14
US-10-140-002-257
; Sequence 257, Application US/10140002
; Patent No. 6725730
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C59
; CURRENT APPLICATION NUMBER: US/10/140,002
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 257
; LENGTH: 1606
; TYPE: DNA
; ORGANISM: Homo Sapien

```

```

US-10-140-002-257

Alignment Scores:
Pred. No.: 148 Length: 1606
Score: 44.00 Matches: 7
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 46.67% Mismatches: 5
Query Match: 46.81% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-68 (1-16) x US-10-140-002-257 (1-1606)
QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValle 15
Db 480 TTCTCTTGGACACACATCACTGATTCATTGATCTGTCACCCCTG 524

RESULT 15
US-09-599-360B-48
; Sequence 48, Application US/09599360B
; Patent No. 6548633
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Bougueleret, L.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal Peptides
; FILE REFERENCE: GENSET.050CP3
; CURRENT APPLICATION NUMBER: US/09/599,360B
; CURRENT FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: 60/113,686
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/141,032
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 09/469,099
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: Patent.pm
; SEQ ID NO 48
; LENGTH: 1622
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..1248
; NAME/KEY: sig_peptide
; LOCATION: 31..135
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 6.30
; OTHER INFORMATION: seq TLLFLFAAPFGLG/EK
; NAME/KEY: polyA_signal
; LOCATION: 1580..1585
; NAME/KEY: polyA_site
; LOCATION: 1607..1622
US-09-599-360B-48

Alignment Scores:
Pred. No.: 150 Length: 1622
Score: 44.00 Matches: 7
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 46.67% Mismatches: 5
Query Match: 46.81% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-68 (1-16) x US-09-599-360B-48 (1-1622)
QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValle 15
Db 493 TTCTCTTGGACACACATCACTGATTCATTGATCTGTCACCCCTG 537

Search completed: November 19, 2004, 00:28:01
Job time : 131.582 secs

```

THIS PAGE BLANK (USPIO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 19, 2004, 22:22:21 ; Search time 73.6 Seconds
(without alignments)

1174.701 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 94

Sequence: 1 FSWDNFCGKDPVIR 16

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3627888 seqs, 2701811610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10030937/runat_16112004_153015_3035/app.query.fasta_1.789
-DB=Published Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-TRANSPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10030937 @CGN 1 1 1223 @runat 16112004 153015 3035
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=0.5 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/PT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/PTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
- 18: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
- 19: /cgn2_6/ptodata/1/pubpna/US11_PUB.seq*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	90.4	546	13	US-10-027-632-207798 Sequence 207798,
2	85	90.4	546	13	US-10-027-632-207799 Sequence 207799,
3	85	90.4	546	13	US-10-027-632-207800 Sequence 207800,
4	85	90.4	546	13	US-10-027-632-207801 Sequence 207801,
5	85	90.4	546	15	US-10-027-632-207798 Sequence 207798,
6	85	90.4	546	15	US-10-027-632-207799 Sequence 207799,
7	85	90.4	546	15	US-10-027-632-207800 Sequence 207800,
8	85	90.4	546	15	US-10-027-632-207801 Sequence 207801,
9	85	90.4	1935	10	US-09-971-392-102 Sequence 102, App
10	85	90.4	2384	9	US-09-822-849A-53 Sequence 53, Appl
11	85	90.4	2436	9	US-09-954-531-380 Sequence 380, App
12	85	90.4	2436	10	US-09-525-978B-81 Sequence 81, Appl
13	85	90.4	2478	15	US-10-170-385-390 Sequence 390, App
14	82	87.2	1983	16	US-10-388-934-167 Sequence 167, App
15	58	61.7	857	13	US-10-027-632-164063 Sequence 164063,
16	58	61.7	857	13	US-10-027-632-164064 Sequence 164064,
17	58	61.7	857	13	US-10-027-632-164065 Sequence 164065,
18	58	61.7	857	15	US-10-027-632-164063 Sequence 164063,
19	58	61.7	857	15	US-10-027-632-164064 Sequence 164064,
20	58	61.7	857	15	US-10-027-632-164065 Sequence 164065,
21	57	60.6	380	10	US-09-764-891-2390 Sequence 2290, App
22	57	60.6	380	16	US-10-264-049-436 Sequence 436, App
23	55	58.5	580	13	US-10-027-632-91348 Sequence 91348, A
24	55	58.5	580	13	US-10-027-632-317712 Sequence 317712,
25	55	58.5	580	15	US-10-027-632-91348 Sequence 91348, A
26	55	58.5	580	15	US-10-027-632-317712 Sequence 317712,
27	50.5	53.7	400660	17	US-10-388-838-68 Sequence 68, Appl
28	50	53.2	2243	16	US-10-108-260A-301 Sequence 301, Appl
29	50	53.2	2367	16	US-10-267-502-81 Sequence 81, Appl
30	50	53.2	2452	15	US-10-094-743-87 Sequence 87, Appl
31	49	52.1	653	13	US-10-027-632-121823 Sequence 121823,
32	49	52.1	653	15	US-10-027-632-121823 Sequence 20544, A
33	49	52.1	1197	16	US-10-424-599-20544 Sequence 78156, A
34	49	52.1	1493	16	US-10-424-599-78156 Sequence 3769, Ap
35	49	52.1	1701	16	US-10-335-977-3769 Sequence 3769, Ap
36	49	52.1	1704	16	US-10-335-977-3768 Sequence 3768, Ap
37	49	52.1	6306	16	US-10-433-794-38 Sequence 38, Appl
38	49	52.1	6629	15	US-10-353-690-35 Sequence 4, Appl
39	49	52.1	6629	15	US-10-172-118-4 Sequence 13, Appl
40	49	52.1	6629	15	US-10-377-035-13 Sequence 339, Appl
41	49	52.1	6629	16	US-10-240-425-339 Sequence 4, Appl
42	49	52.1	6629	16	US-10-342-887-4 Sequence 11, Appl
43	49	52.1	7676	16	US-10-042-865-11 Sequence 11, Appl
44	49	52.1	7826	15	US-10-377-035-11 Sequence 32, Appl
45	49	52.1	7862	15	US-10-288-798-32

ALIGNMENTS

RESULT 1
US-10-027-632-207798
; Sequence 207798, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; TITLE OF INVENTION: 108827.129
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23

```
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207798
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-207798
```

```
Alignment Scores:
Pred. No.:          9.05e-06      Length:      546
Score:             85.00         Matches:    15
Percent Similarity: 93.75%       Conservative: 0
Best Local Similarity: 93.75%    Mismatches:  1
Query Match:       90.43%        Indels:     0
DB:                13           Gaps:         0
```

US-10-030-937-68 (1-16) x US-10-027-632-207798 (1-546)

```
QY      1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
        |||||||
Db       276 TTTTCTGGGATACTGTGATGAAGGGAAGGACCCCTGCGGTGATCAGA 323
```

RESULT 2

```
US-10-027-632-207799
; Sequence 207799, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207799
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-207799
```

```
Alignment Scores:
Pred. No.:          9.05e-06      Length:      546
Score:             85.00         Matches:    15
Percent Similarity: 93.75%       Conservative: 0
Best Local Similarity: 93.75%    Mismatches:  1
Query Match:       90.43%        Indels:     0
DB:                13           Gaps:         0
```

US-10-030-937-68 (1-16) x US-10-027-632-207799 (1-546)

```
QY      1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
        |||||||
Db       276 TTTTCTGGGATACTGTGATGAAGGGAAGGACCCCTGCGGTGATCAGA 323
```

RESULT 3

```
US-10-027-632-207800
; Sequence 207800, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207800
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-207800
```

Alignment Scores:

```
Pred. No.:          9.05e-06      Length:      546
Score:             85.00         Matches:    15
Percent Similarity: 93.75%       Conservative: 0
Best Local Similarity: 93.75%    Mismatches:  1
Query Match:       90.43%        Indels:     0
DB:                13           Gaps:         0
```

US-10-030-937-68 (1-16) x US-10-027-632-207800 (1-546)

```
QY      1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
        |||||||
Db       276 TTTTCTGGGATACTGTGATGAAGGGAAGGACCCCTGCGGTGATCAGA 323
```

RESULT 4

```
US-10-027-632-207801
; Sequence 207801, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 207801
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-207801

Alignment Scores:
Pred. No.: 9.05e-06 Length: 546
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 13 Gaps: 0

US-10-030-937-68 (1-16) x US-10-027-632-207801 (1-546)

Qy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
|||
Db 276 TTTTCTGGGATAACTGTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 323

RESULT 5

US-10-027-632-207798
; Sequence 207798, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; POLYMORPHISMS IN THE HUMAN GENOME

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 207798

; LENGTH: 546

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-207798

Alignment Scores:
Pred. No.: 9.05e-06 Length: 546
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 15 Gaps: 0

US-10-030-937-68 (1-16) x US-10-027-632-207798 (1-546)

Qy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
|||
Db 276 TTTTCTGGGATAACTGTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 323

RESULT 6

US-10-027-632-207799
; Sequence 207799, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 207799

; LENGTH: 546

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-207799

Alignment Scores:
Pred. No.: 9.05e-06 Length: 546
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 15 Gaps: 0

US-10-030-937-68 (1-16) x US-10-027-632-207799 (1-546)

Qy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
|||
Db 276 TTTTCTGGGATAACTGTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 323

RESULT 7

US-10-027-632-207800
; Sequence 207800, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; POLYMORPHISMS IN THE HUMAN GENOME

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 207800

; LENGTH: 546

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-207800

Alignment Scores:
Pred. No.: 9.05e-06 Length: 546
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 15 Gaps: 0

US-10-030-937-68 (1-16) x US-10-027-632-207800 (1-546)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 276 TTTTCTGGGATACTGTGATGAAGGAGGACCTCGGTGATCAGA 323

RESULT 8

US-10-027-632-207801
; Sequence 207801, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207801
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-207801

Alignment Scores:
Pred. No.: 9.05e-06 Length: 546
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 15 Gaps: 0

US-10-030-937-68 (1-16) x US-10-027-632-207801 (1-546)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 276 TTTTCTGGGATACTGTGATGAAGGAGGACCTCGGTGATCAGA 323

RESULT 9

US-09-971-392-102
; Sequence 102, Application US/09971392
; Publication No. US20030134283A1
; GENERAL INFORMATION:
; APPLICANT: Peterson, David P.
; APPLICANT: Pearson, Cecelia I.
; APPLICANT: Cocks, Benjamin G.
; TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
; FILE REFERENCE: PA-0029 US

; CURRENT APPLICATION NUMBER: US/09/971,392
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 60/237,652
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PERL Program
; SEQ ID NO 102
; LENGTH: 1935
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Template ID: 977615.8
US-09-971-392-102

Alignment Scores:
Pred. No.: 3.7e-05 Length: 1935
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 10 Gaps: 0

US-10-030-937-68 (1-16) x US-09-971-392-102 (1-1935)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 201 TTTTCTGGGATACTGTGATGAAGGAGGACCTCGGTGATCAGA 248

RESULT 10

US-09-822-849A-53
; Sequence 53, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 2384
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-53

Alignment Scores:
Pred. No.: 4.67e-05 Length: 2384
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-68 (1-16) x US-09-822-849A-53 (1-2384)

QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 112 TTTTCTGGGATACTGTGATGAAGGAGGACCTCGGTGATCAGA 159

RESULT 11

US-09-954-531-380

```

; Sequence 380, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 380
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-380

Alignment Scores:
Pred. No.: 4.78e-05 Length: 2436
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-68 (1-16) x US-09-954-531-380 (1-2436)
QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 158 TTTTCTGGGATAACTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 205

RESULT 12
; Sequence 81, Application US/09525978B
; Publication No. US20030049722A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Caras, Ingrid W.
; APPLICANT: Hevezi, Peter
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING MACROPHAGE DEVELOPMENT
; TITLE OF INVENTION: RELATED DISORDERS, COMPOSITIONS, AND METHODS OF
; TITLE OF INVENTION: SCREENING FOR MACROPHAGE DEVELOPMENT MODULATORS
; FILE REFERENCE: A-67413-1/DJB/JUD
; CURRENT APPLICATION NUMBER: US/09/525,978B
; CURRENT FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: USSN 60/124,530
; PRIOR FILING DATE: 1999-03-15
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-525-978B-81

Alignment Scores:
Pred. No.: 4.78e-05 Length: 2436
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 10 Gaps: 0

US-10-030-937-68 (1-16) x US-09-525-978B-81 (1-2436)
QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 158 TTTTCTGGGATAACTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 205

RESULT 14
; Sequence 167, Application US/10388934
; Publication No. US20040005547A1
; GENERAL INFORMATION:
; APPLICANT: Boess, Franziska
; APPLICANT: Suter-Dick, Laura
; APPLICANT: Wolf, Detlef
; TITLE OF INVENTION: BIOMARKERS AND EXPRESSION PROFILES FOR TOXICOLOGY
; FILE REFERENCE: 21199
; CURRENT APPLICATION NUMBER: US/10/388,934
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 02005336.9
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 02015657.6
; PRIOR FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 862
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 167
; LENGTH: 1983
; TYPE: DNA
US-10-388-934-167

Alignment Scores:
Pred. No.: 4.88e-05 Length: 2478
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 15 Gaps: 0

US-10-030-937-68 (1-16) x US-10-388-934 (1-2478)
QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 195 TTTTCTGGGATAACTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 242

RESULT 14
; Sequence 167, Application US/10388934
; Publication No. US20040005547A1
; GENERAL INFORMATION:
; APPLICANT: Boess, Franziska
; APPLICANT: Suter-Dick, Laura
; APPLICANT: Wolf, Detlef
; TITLE OF INVENTION: BIOMARKERS AND EXPRESSION PROFILES FOR TOXICOLOGY
; FILE REFERENCE: 21199
; CURRENT APPLICATION NUMBER: US/10/388,934
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 02005336.9
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 02015657.6
; PRIOR FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 862
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 167
; LENGTH: 1983
; TYPE: DNA
US-10-388-934-167

Alignment Scores:
Pred. No.: 4.88e-05 Length: 2478
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 15 Gaps: 0

US-10-030-937-68 (1-16) x US-10-388-934 (1-2478)
QY 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db 195 TTTTCTGGGATAACTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 242

```

```
; ORGANISM: Rattus sp.
US-10-388-934-167

Alignment Scores:
Pred. No.:      0.000131      Length:      1983
Score:          82.00         Matches:      14
Percent Similarity: 93.75%    Conservative: 1
Best Local Similarity: 87.50%  Mismatches:   1
Query Match:    87.23%       Indels:       0
DB:             16           Gaps:         0

US-10-030-937-68 (1-16) x US-10-388-934-167 (1-1983)

QY      1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db      129 TTCTCTGGGATACTGTGATGAAGGAAGGACCCCTGTCAGTGATCAAA 176

RESULT 15
US-10-027-632-164063/c
; Sequence 164063, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 164063
; LENGTH: 857
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-164063

Alignment Scores:
Pred. No.:      1.01      Length:      857
Score:          58.00     Matches:      10
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 62.50% Mismatches:   4
Query Match:    61.70%       Indels:       0
DB:             13         Gaps:         0

US-10-030-937-68 (1-16) x US-10-027-632-164063 (1-857)

QY      1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
Db      216 TTTCCTGGGATACTGGACAAAGGAAGGACCCCTGTCAGTGATCAAA 169

Search completed: November 19, 2004, 03:29:41
Job time : 75.6 secs
```

Fri Nov 19 14:12:30 2004

us-10-030-937-68.p2n.rst

```

us-10-030-937-68.p2n.rst

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame plus_p2n model
Run on: November 18, 2004, 19:16:51 ; Search time 514.773 Seconds
(without alignments)
1132.607 Million cell updates/sec

Title:
Perfect score: 94
Sequence: 1 FSWDNCFEKDPVAVR 16

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 1821985908 residues
Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool_p/US10030937/runat_16112004_153014_2925/app.query.fasta_1.789
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937 -CGCN 1 1 9321 @runat_16112004_153014_2925 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description
-----
1 85 90.4 352 7 T28061 T28061 EST25975 Hu
2 85 90.4 383 6 CD695415 CD695415 EST11938
3 85 90.4 434 6 CD710184 CD710184 EST26711
4 85 90.4 438 2 BE041669 BE041669 hc33e04.x
5 85 90.4 450 7 H00793 H00793 Yj30e08.r1
6 85 90.4 455 6 CD710451 CD710451 EST26978
7 85 90.4 476 4 BM147068 BM147068 TCAP1010
8 85 90.4 517 1 AA306060 AA306060 EST177045
9 85 90.4 533 1 A1793102 A1793102 on44c06.y

10 85 90.4 554 4 BG830178 BG830178
11 85 90.4 559 1 AU280628 AU280628
12 85 90.4 600 5 BX473154 BX473154
13 85 90.4 601 5 BX506263 BX506263
14 85 90.4 603 2 BF955424 BF955424
15 85 90.4 611 4 BG623044 BG623044
16 85 90.4 633 4 BG773758 BG773758
17 85 90.4 643 1 AU135438 AU135438
18 85 90.4 646 6 CD703135 CD703135
19 85 90.4 651 4 BM723945 BM723945
20 85 90.4 660 4 BG470369 BG470369
21 85 90.4 691 2 AW953554 AW953554
22 85 90.4 698 2 BF509172 BF509172
23 85 90.4 711 4 BG479322 BG479322
24 85 90.4 726 4 BG762599 BG762599
25 85 90.4 733 4 BI856212 BI856212
26 85 90.4 760 4 BG770447 BG770447
27 85 90.4 767 4 BI161208 BI161208
28 85 90.4 784 4 BG323734 BG323734
29 85 90.4 784 4 BI838554 BI838554
30 85 90.4 813 4 BG913328 BG913328
31 85 90.4 818 1 AL552056 AL552056
32 85 90.4 842 4 BI091220 BI091220
33 85 90.4 847 5 BI220522 BI220522
34 85 90.4 855 1 A1547002 A1547002
35 85 90.4 858 4 BI820051 BI820051
36 85 90.4 863 4 BG478588 BG478588
37 85 90.4 870 5 BX404478 BX404478
38 85 90.4 885 6 CA453926 CA453926
39 85 90.4 895 5 BQ676659 BQ676659
40 85 90.4 900 6 CB231270 CB231270
41 85 90.4 910 1 AL560604 AL560604
42 85 90.4 911 2 BF127825 BF127825
43 85 90.4 912 1 AL543858 AL543858
44 85 90.4 937 5 BU151364 BU151364
45 85 90.4 948 5 BU556606 BU556606

ALIGNMENTS

T28061 352 bp mRNA linear EST 06-SEP-1995
EST25975 Human Brain Homo sapiens cDNA 5' end similar to activator
protein GM2 (HT:2720), mRNA sequence.

T28061 GI:610159
T28061.1 Homo sapiens (human)
EST.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N., Kirkness,E.P., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fritchman,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,P.S.,
Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
Merrick,J.M., Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T.,
Pellegriano,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A.,
Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,
Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
Meissner,P., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
Initial Assessment of Human Gene Diversity and Expression Patterns
Based Upon 83 Million Basepairs of cDNA Sequence
Nature 377, 3-174 (1995)

```

MEDLINE

96026280
7566098
Contact: Venter, JC
The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@db.tigr.org
For clone availability, additional sequence and expression
information related to this EST, please contact the TIGR Database
(tdbinfo@db.tigr.org)
Seq primer: M13 Reverse.
Location/Qualifiers
1..352
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="ATCC (inhost):101009"
/db_xref="taxon:9606"
/clone_lib="Human Brain"
/note="Organ: brain"

FEATURES

source
Location/Qualifiers
1..352
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="ATCC (inhost):101009"
/db_xref="taxon:9606"
/clone_lib="Human Brain"
/note="Organ: brain"

ORIGIN

Alignment Scores:
Pred. No.: 0.000382 Length: 352
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 7 Gaps: 0

US-10-030-937-68 (1-16) x T28061 (1-352)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16

Db 95 TTTTCTGGGATAACTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 142

RESULT 2

LOCUS CD695415 383 bp mRNA linear EST 25-JUN-2003
DEFINITION EST11938 human nasopharynx Homo sapiens cDNA, mRNA sequence.
ACCESSION CD695415
VERSION CD695415.1 GI:32220943
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Liu, X.-Q., Zhou, Y., Zhang, L.-J., Xu, H., Chen, H.-K., Pan, Z.-G. and
Zeng, Y.-X.
TITLE Transcriptional Gene Expression Profile of Human Nasopharynx
JOURNAL Unpublished (2003)
COMMENT Contact: Yixin Zeng
Cancer Center
Sun Yat-sen University
651 Dongfeng Road East, Guangzhou 510060, China
Tel: 86-1380-9770-743
Fax: 86-20-8775-4506
Email: yxzeng@gzsums.edu.cn.
Location/Qualifiers
1...383
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clone_lib="human nasopharynx"
/note="ESTs generated from a normal nasopharynx cDNA
library from southern Chinese"

ORIGIN

Alignment Scores:
Pred. No.: 0.000421 Length: 383
Score: 85.00 Matches: 15

Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-68 (1-16) x CD695415 (1-383)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16

Db 161 TTTTCTGGGATAACTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 208

RESULT 3

LOCUS CD710184 434 bp mRNA linear EST 25-JUN-2003
DEFINITION EST26711 human nasopharynx Homo sapiens cDNA, mRNA sequence.
ACCESSION CD710184
VERSION CD710184.1 GI:32240814
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Liu, X.-Q., Zhou, Y., Zhang, L.-J., Xu, H., Chen, H.-K., Pan, Z.-G. and
Zeng, Y.-X.
TITLE Transcriptional Gene Expression Profile of Human Nasopharynx
JOURNAL Unpublished (2003)
COMMENT Contact: Yixin Zeng
Cancer Center
Sun Yat-sen University
651 Dongfeng Road East, Guangzhou 510060, China
Tel: 86-1380-9770-743
Fax: 86-20-8775-4506
Email: yxzeng@gzsums.edu.cn.
Location/Qualifiers
1..434
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clone_lib="human nasopharynx"
/note="ESTs generated from a normal nasopharynx cDNA
library from southern Chinese"

FEATURES

source
Location/Qualifiers
1..434
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clone_lib="human nasopharynx"
/note="ESTs generated from a normal nasopharynx cDNA
library from southern Chinese"

ORIGIN

Alignment Scores:
Pred. No.: 0.000486 Length: 434
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-68 (1-16) x CD710184 (1-434)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16

Db 217 TTTTCTGGGATAACTGTGATGAAGGAGGACCCCTGCGGTGATCAGA 264

FEATURES

source
Location/Qualifiers
1...383
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clone_lib="human nasopharynx"
/note="ESTs generated from a normal nasopharynx cDNA
library from southern Chinese"

ORIGIN

Alignment Scores:
Pred. No.: 0.000421 Length: 383
Score: 85.00 Matches: 15

BE041669 438 bp mRNA linear EST 08-JUN-2000
LOCUS ho33e04.x1 NCI CGAP Lu24 Homo sapiens cDNA clone IMAGE:3039198 3,
DEFINITION similar to gb:X62078 GANGLIOSIDE GM2 ACTIVATOR PRECURSOR (HUMAN);,
mRNA sequence.
ACCESSION BE041669
VERSION BE041669.1 GI:9358722
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 438)

Fri Nov 19 14:12:30 2004

us-10-030-937-68.p2n.rst

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
 Possible reversed clone: polyT not found
 Seq primer: -40UP from Gibco
 High quality sequence stop: 392.
 Location/Qualifiers
 1..438
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3039198"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Lu24"
 /notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI CGAP Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN
 Alignment Scores: Length: 438
 Pred. No.: 0.000491 Matches: 15
 Score: 85.00 Conservative: 0
 Percent Similarity: 93.75% Mismatches: 1
 Best Local Similarity: 93.75% Indels: 0
 Query Match: 90.43% Gaps: 0
 DB: 2

US-10-030-937-68 (1-16) x BR041669 (1-438)
 Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 |||||
 Db 150 TTTTCTGGGATAACTGTGATGAAGGAGGACCTCGCGGTGATCAGA 197
 |||||

RESULT 5
 H00793 450 bp mRNA linear EST 19-JUN-1995
 LOCUS YJ30e08.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
 DEFINITION IMAGE:150278 5' similar to gb:X62078 GANGLIOSIDE GM2 ACTIVATOR
 KEYWORDS PRECURSOR (HUMAN);, mRNA sequence.
 H00793
 ACCESSION H00793.1 GI:863726
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE EST.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 450)
 Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Huitman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
 Wilson,R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
 Possible reversed clone: polyT not found
 Seq primer: -40UP from Gibco
 High quality sequence stop: 392.
 Location/Qualifiers
 1..438
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3039198"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Lu24"
 /notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI CGAP Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN
 Alignment Scores: Length: 438
 Pred. No.: 0.000491 Matches: 15
 Score: 85.00 Conservative: 0
 Percent Similarity: 93.75% Mismatches: 1
 Best Local Similarity: 93.75% Indels: 0
 Query Match: 90.43% Gaps: 0
 DB: 2

US-10-030-937-68 (1-16) x BR041669 (1-438)
 Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 |||||
 Db 150 TTTTCTGGGATAACTGTGATGAAGGAGGACCTCGCGGTGATCAGA 197
 |||||

RESULT 5
 H00793 450 bp mRNA linear EST 19-JUN-1995
 LOCUS YJ30e08.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
 DEFINITION IMAGE:150278 5' similar to gb:X62078 GANGLIOSIDE GM2 ACTIVATOR
 KEYWORDS PRECURSOR (HUMAN);, mRNA sequence.
 H00793
 ACCESSION H00793.1 GI:863726
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE EST.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 450)
 Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Huitman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
 Wilson,R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

Insert Size: 1053
 High quality sequence stops: 298
 Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1053 Std Error: 0.00
 Seq primer: M13RP1

High quality sequence stop: 298.
 Location/Qualifiers

1..450
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:562247"
 /db_xref="taxon:9606"
 /clone="IMAGE:150278"
 /sex="Female"
 /dev_stage="placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares placenta Nb2HP"
 /note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGAAGAAATCGCGCGGAGGAAATTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Alignment Scores: Length: 450
 Pred. No.: 0.000506 Matches: 15
 Score: 85.00 Conservative: 0
 Percent Similarity: 93.75% Mismatches: 1
 Best Local Similarity: 93.75% Indels: 0
 Query Match: 90.43% Gaps: 0
 DB: 7

US-10-030-937-68 (1-16) x H00793 (1-450)

Qy 1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
 |||||
 Db 177 TTTTCTGGGATAACTGTGATGAAGGAGGACCTCGCGGTGATCAGA 224
 |||||

RESULT 6

CD710451 465 bp mRNA linear EST 25-JUN-2003
 LOCUS EST26978 human nasopharynx Homo sapiens cDNA, mRNA sequence.

DEFINITION CD710451

ACCESSION CD710451.1 GI:32241081

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 465)

Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and

Zeng,Y.-X.

Transcriptional Gene Expression Profile of Human Nasopharynx

Unpublished (2003)

CONTACT: Yixin Zeng

Cancer Center

Sun Yat-sen University

651 Dongfeng Road East, Guangzhou 510060, China

Tel: 86-1380-9770-743

Fax: 86-20-8775-4506

Email: yxzeng@zsuums.edu.cn

Location/Qualifiers

FEATURES

```

source
1. .465
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clone_lib="human nasopharynx"
/notes="ESTs generated from a normal nasopharynx cDNA
library from southern Chinese"

ORIGIN
Alignment Scores:
Pred. No.: 0.000526 Length: 465
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservatives: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-68 (1-16) x CD710451 (1-465)

QY 1 PheSerTrpAspAsnCysPheGluGlyeAspProAlaValIleArg 16
|||||
Db 146 TTTTCTGGGATAACTGTGATGAAGGAGACCCCTGGCGTGATCAGA 193

RESULT 7
BM147068
LOCUS
DEFINITION
TCAAP1Q10171 Pediatric acute myelogenous leukemia cell (FAB M1)
Baylor-HGSC project=TCAA Homo sapiens cDNA clone TCAAP1017, mRNA
sequence.
ACCESSION
BM147068
VERSION
BM147068
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 476)
AUTHORS
Wei,Y., Teang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman,F.R. Jr.,
Gunarathne,P.H., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F.
Pediatric Leukemia cDNA Sequencing Project (2001)
Unpublished (2001)
Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@txccc.org
Seq primer: M13 primer.
Location/Qualifiers
1. .476
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TCAAP1017"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="myeloid cell"
/dev_stage="pediatric 6 years"
/lab_host="DH10B"
/clone_lib="Pediatric acute myelogenous leukemia cell (FAB
M1) Baylor-HGSC project=TCAA"
/notes="Vector: lambda pSB; Site 1: BamHI; Site 2: EcoRI;
First strand cDNA was primed with an anchored
xhoI-oligo(dT) primer [5'GGAGACTCGAGCGCCGACGAGAG(T)VN
3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand
was primed with a BamHI-dc primer
[5'AGAGAGTCGATCCGCGCCCAATAATAAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and SalI sites of
lambda pSB vector. Library went through one round of
normalization. Library was constructed by Wei Yu at RIKEN

```

of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T, Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M, Schneider C, Hayashizaki Y, High efficiency selection of full-length cDNA by improved biotinylated cap trapper., DNA Res 4: 1, 61-6, Feb 28, 1997)"

ORIGIN

```

Alignment Scores:
Pred. No.: 0.00054 Length: 476
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservatives: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-68 (1-16) x BM147068 (1-476)

QY 1 PheSerTrpAspAsnCysPheGluGlyeAspProAlaValIleArg 16
|||||
Db 106 TTTTCTGGGATAACTGTGATGAAGGAGACCCCTGGCGTGATCAGA 153

RESULT 8
AA306060
LOCUS
DEFINITION
EST177045 Jurkat T-cells VI Homo sapiens cDNA 5' end similar to
similar to activator protein GM2, mRNA sequence.
ACCESSION
AA306060
VERSION
AA306060.1 GI:1958388
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 517)
AUTHORS
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palances,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudke,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meisner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
PUBMED
7566098
Other ESTs: THC194805
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
Location/Qualifiers
1. .517
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="ATCC (inhost):160212"

```

FEATURES

source

/db_xref="taxon:9606"
 /cell_type="T-lymphocyte"
 /clone_lib="Jurkat T-cells VI"
 /notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI"

ORIGIN

Alignment Scores:
 Pred. No.: 0.000594 Length: 517
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 1 Gaps: 0

US-10-030-937-68 (1-16) x AA306060 (1-517)

QY 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
 |||||
 Db 247 TTTTCTGGGATAACTGTGATGAAGGAGGACCTCGGTGATCAGA 294
 |||||

RESULT 9
 AI793102

LOCUS AI793102 533 bp mRNA linear EST 13-DEC-1999
 DEFINITION on44c06.y5 NCI_CGAP Co8 Homo sapiens cDNA clone IMAGE:1559530 5'
 similar to gb:X62078 GANGLIOSIDE GM2 ACTIVATOR PRECURSOR (HUMAN) ;
 mRNA sequence.

ACCESSION AI793102.1 GI:5340818
 VERSION AI793102.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 533)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Other ESTs: on44c06.x5
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moshaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: see original entry for original citation
 information
 This 5' resequenced clone has no previous 5' data to verify this
 new read against
 Insert Length: 1207 Std Error: 0.00
 Seg primer: -40RP from Gibco
 High quality sequence stop: 455.
 Location/Qualifiers
 1..533
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1559530"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Co8"
 /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 colon adenocarcinoma, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned

FEATURES

source

into the Not I and Eco RI sites of the modified pT7T3
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Alignment Scores:
 Pred. No.: 0.000615 Length: 533
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 1 Gaps: 0

US-10-030-937-68 (1-16) x AI793102 (1-533)

QY 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
 |||||
 Db 110 TTTTCTGGGATAACTGTGATGAAGGAGGACCTCGGTGATCAGA 157
 |||||

RESULT 10

AI793102

LOCUS BG830178 554 bp mRNA linear EST 22-MAY-2001
 DEFINITION 602764754F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4906909 5',
 mRNA sequence.

ACCESSION BG830178
 VERSION BG830178.1 GI:14177765
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC

FEATURES

source

1..554
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4906909"
 /tissue_type="epithelioid carcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_42"
 /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1-8kb. Library constructed by Ling
 Hong in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library. |"

ORIGIN

Alignment Scores:
 Pred. No.: 0.000643 Length: 554
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 4 Gaps: 0

```

US-10-030-937-68 (1-16) x BG830178 (1-554)

QY      1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
      |||
Db      173 TTTTCTGGGATAACTGTGTGTAAGGAGGACCTCGCGTGATCAGA 220

RESULT 11
LOCUS   AU280628
DEFINITION AU280628 NCRRM1 Homo sapiens cDNA clone NCRRM1000016 5', mRNA
sequence.
ACCESSION AU280628
VERSION   AU280628.1 GI:28299855
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 559)
AUTHORS   Imabayashi,H., Mori,T., Gojo,S., Kiyono,T., Sugiyama,T., Irie,R.,
          Isogai,T., Hata,J., Tomoya,Y. and Umezawa,A.
TITLE      Redifferentiation of dedifferentiated chondrocytes and
          chondrogenesis of human bone marrow stromal cells via chondrosphere
          formation with expression profiling by large-scale cDNA analysis
JOURNAL   Exp. Cell Res. 288 (1), 35-50 (2003)
MEDLINE   22760698
PUBMED    12878157
COMMENT    Contact: Takao Isogai
          Genomics Laboratory
          Helix Research Institute
          1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
          Tel: 81-438-52-3975
          Fax: 81-438-52-3986
          Email: genomics@hri.co.jp
          HRI human cDNA Project, Sugiyama,T.; Wakamatsu,A.; Irie,R.;
          Umezawa,A.; Fukuma,M.; Kusakari,S.; Hata,J.; Ishii,S.; Yamamoto,J.;
          Isono,Y.; Saito,K.; Nakamura,Y.; Masubo,Y.; Nagai,K.; Isogai,T.
          HRI human cDNA project: cDNA library construction & 5'-end one
          pass sequencing: Helix Research Institute.
FEATURES
source   1..559
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="NCRRM1000016"
          /cell_type="embryonal carcinoma"
          /clone_lib="NCRRM1"
          /note="Vector: pME18SFL3; mRNA from uninduced embryonal
          carcinoma"

ORIGIN
Alignment Scores:
Pred. No.: 0.00065 Length: 559
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 1 Gaps: 0

US-10-030-937-68 (1-16) x AU280628 (1-559)

QY      1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
      |||
Db      132 TTTTCTGGGATAACTGTGTGTAAGGAGGACCTCGCGTGATCAGA 179

RESULT 12
LOCUS   BX473154
DEFINITION BX473154 DKFZp686H18154 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
sequence.
ACCESSION BX473154
VERSION   BX473154.1 GI:31667405
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 600)
AUTHORS   Koerber,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
          Fobo,G., Han,M. and Wiemann,S.
          EST (Koerber,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., et al.)
          Unpublished (2003)
          Contact: MIPS
          MIPS
          Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
          This is the 5' sequence of the clone insert
          Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
          Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
          sequenced by BMFZ (Biomedical Research Center at the Heinrich-
          Heine-University, Duesseldorf/Germany) within the cDNA sequencing
          consortium of the German Genome Project. No sl sequence available.
          This clone (DKFZp686H18154) is available at the RZPD in Berlin.
          Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
          Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
          Location/Qualifiers
          1..600
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="DKFZp686H18154"
          /dev_stage="adult"
          /lab_host="DH10B"
          /clone_lib="686 (synonym: hlcc3)"
          /note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
          cDNA-collection"

ORIGIN
Alignment Scores:
Pred. No.: 0.000705 Length: 600
Score: 85.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 90.43% Indels: 0
DB: 5 Gaps: 0

US-10-030-937-68 (1-16) x BX473154 (1-600)

QY      1 PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg 16
      |||
Db      243 TTTTCTGGGATAACTGTGTGTAAGGAGGACCTCGCGTGATCAGA 290

RESULT 13
LOCUS   BX506263
DEFINITION BX506263 DKFZp686F03234 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
sequence.
ACCESSION BX506263
VERSION   BX506263.1 GI:32037321
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 601)
AUTHORS   Ottenwaelder,B., Obermaier,B., Deutschbauer,S., Mewes,H.W.,
          Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
          EST (Ottenwaelder,B., Obermaier,B., Deutschbauer,S., Mewes,H.W.,
          et al.)
          Unpublished (2003)
          Contact: MIPS
          MIPS
          Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
          This is the 5' sequence of the clone insert
          Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
          Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

```

sequenced by MediGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No si sequence available.

This clone (DKFP686F03234) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin- Charlottenburg, GERMANY; Email: cloneerzpd.de.

FEATURES

Location/Qualifiers
 1. .601
 /organism="Homo sapiens"
 /mol_type="mrna"
 /db_xref="taxon:9606"
 /clone="DKFP686F03234"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="686 (synonym: hlcc3)"
 /note="Vector: pTriplex2; Site 1: SfiIA; Site 2: SfiIB; cDNA-collection"

ORIGIN

Alignment Scores:
 Pred. No.: 0.000707 Length: 601
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 5 Gaps: 0

US-10-030-937-68 (1-16) x BX506263 (1-601)

Qy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
 |||||
 Db 183 TTTTCCTGGGATAACTGTGATGAAGGAAGACCCCTGCGGTGATCAGA 230

RESULT 14

BF955424
 LOCUS M84-NN1196-171100-001-a07 NN1196 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF955424
 VERSION BF955424.1 GI:12372686

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1. (bases 1 to 603)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

TITLE

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

20202663

MEDLINE

10737800

PUBMED

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR4&t2=MR4-NN1196-171100-001-a07&t3=2000-11-17&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 34

High quality sequence stop: 583.

FEATURES

Location/Qualifiers

1. .603

/organism="Homo sapiens"
 /mol_type="mrna"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="NN1196"

/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Alignment Scores:
 Pred. No.: 0.000709 Length: 603
 Score: 85.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 90.43% Indels: 0
 DB: 2 Gaps: 0

US-10-030-937-68 (1-16) x BF955424 (1-603)

Qy 1 PheSerTrpAspAsnCysPheGluGlyAspProAlaValIleArg 16
 |||||
 Db 205 TTTTCCTGGGATAACTGTGATGAAGGAAGACCCCTGCGGTGATCAGA 252

RESULT 15

BG623044
 LOCUS 602647926F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4769238 5', mRNA sequence.
 ACCESSION BG623044
 VERSION BG623044.1 GI:13674415

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1. (bases 1 to 611)

NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:

http://image.llnl.gov

Plate: L1CM1634 row: d column: 07

High quality sequence stop: 611.

FEATURES

Location/Qualifiers

1. .611

source

/organism="Homo sapiens"

/mol_type="mrna"

/db_xref="taxon:9606"

/clone="IMAGE:4769238"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH MGC 79"

/note="Organ: placenta; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGCC-3' and 3' adaptor sequence:

5'-ATTCTAGAGCGGAGCGGCGGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.:	0.00072	Length:	611
Score:	85.00	Matches:	15
Percent Similarity:	93.75%	Conservative:	0
Best Local Similarity:	93.75%	Mismatches:	1
Query Match:	90.43%	Indels:	0
DB:	4	Gaps:	0

US-10-030-937-68 (1-16) x BG623044 (1-611)

Qy	1	PheSerTrpAspAsnCysPheGluGlyLysAspProAlaValIleArg	16
Db	199	TTTTCTGGATAACTGTGATGAAGGAGGACCTCGGTGATCAGA	246

Search completed: November 19, 2004, 00:22:24
Job time : 516.773 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 06:12:35 ; Search time 4.65196 Seconds
(without alignments)
1233.819 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 94

Sequence: 1 FSWDNCFEGKPAVIR 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 40 summaries

Database : A Geneseq_23Sep04:*

- 1: geneseqp1980s:*
- 2: geneseqp1980s:*
- 3: geneseqp2000s:*
- 4: geneseqp2000s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94	100.0	16	4 AAB31926	Aab31926 Amino aci
2	94	100.0	193	4 AAB31897	Aab31897 Amino aci
3	85	90.4	178	4 AAB31898	Aab31898 Amino aci
4	85	90.4	178	5 ABG31346	Abg31346 Non-glyco
5	85	90.4	189	4 AAB31900	Aab31900 Amino aci
6	85	90.4	193	4 AAB31901	Aab31901 Amino aci
7	85	90.4	193	4 AAB31902	Aab31902 Amino aci
8	85	90.4	193	4 AAB31904	Aab31904 Amino aci
9	85	90.4	193	4 AAB31896	Aab31896 Amino aci
10	85	90.4	193	4 AAB31928	Aab31928 Amino aci
11	85	90.4	193	4 AAB31903	Aab31903 Amino aci
12	85	90.4	193	4 ABG00720	Abg00720 Novel hum
13	85	90.4	193	5 ABG31345	Abg31345 Human GM2
14	85	90.4	193	5 ABP65212	Abp65212 Hypoxia-r
15	85	90.4	193	7 ADN95858	Adn95858 Human BEC
16	85	90.4	193	8 ADN03620	Adn03620 Antipsori
17	85	90.4	193	8 ADQ17712	Adq17712 Human sof
18	85	90.4	200	4 AAB31899	Aab31899 Amino aci
19	82	87.2	199	2 AAW10656	Aaw10656 Rat GM2 a
20	57	60.6	126	4 AAM96319	Aam96319 Human rev
21	57	60.6	191	5 ABP41479	Abp41479 Human ova
22	46	48.9	211	7 ADD30064	Add30064 Plant yie
23	46	48.9	211	8 ADI44383	Adi44383 Plant tra
24	46	48.9	575	4 AAG84972	Aag84972 Shrimp wh
25	46	48.9	588	6 ABU49465	Abu49465 Protein e

RESULT 1
AAB31926
ID AAB31926 standard; peptide; 16 AA.

XX AC AAB31926;

XX DT 15-MAY-2001 (first entry)

XX DE Amino acid sequence of a peptide fragment of a human protein.

XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX OS Homo sapiens.

XX FN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX FR 15-JUL-1999; 99FR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX DR WPI; 2001-159475/16.

XX PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.

XX FS Claim 24; Page 108; 209pp; French.

XX CC The present sequence represents a peptide fragment of a human protein,
CC which is used in the method of the invention. The specification describes
CC a method which uses at least one polypeptide or polynucleotide sequence
CC belonging to the perlecan, precursor of the retinol-binding plasma
CC protein, precursor of the ganglioside GM2 activator, calgranulin B or
CC saposin B protein families. The method is used for detecting, preventing
CC or treating a degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis, and
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and

CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 16 AA;

Query Match 100.0%; Score 94; DB 4; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1.9e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDAVIR 16
 |||||
 DB 1 FSWDNCFEKGKDAVIR 16

RESULT 2
 AAB31897
 ID AAB31897 standard; protein; 193 AA.

AC AAB31897;

DT 15-MAY-2001 (first entry)

DE Amino acid sequence of a mutant ganglioside GM2 activator protein.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX Homo sapiens.

XX WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INMR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX Claim 25; Page 159-160; 209pp; French.

XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 193 AA;

Query Match 100.0%; Score 94; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDAVIR 16
 |||||
 DB 34 FSWDNCFEKGKDAVIR 49

RESULT 3
 AAB31898
 ID AAB31898 standard; protein; 178 AA.

AC AAB31898;

DT 15-MAY-2001 (first entry)

DE Amino acid sequence of a human protein.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX Homo sapiens.

XX WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INMR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX Claim 1; Page 160; 209pp; French.

XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 178 AA;

Query Match 90.4%; Score 85; DB 4; Length 178;
 Best Local Similarity 93.8%; Pred. No. 7.9e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDAVIR 16
 |||||
 DB 19 FSWDNCDEGKDAVIR 34

RESULT 4
 ABG31346
 ID ABG31346 standard; protein; 178 AA.
 XX AC
 XX AC ABG31346;
 XX DT 15-NOV-2002 (first entry)
 XX DE Non-glycosylated human GM2 activator protein containing His6-tag.
 XX KW Human; GM2 activator protein; ganglioside; platelet activating factor;
 KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
 KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
 KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
 KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
 KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
 KW antiinflammatory; antiasthmatic; antiarthritic; antirheumatic;
 KW antiallergic; hepatotropic; nephrotropic; immunosuppressive;
 KW tranquilizer.
 XX OS Homo sapiens.
 OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Binding-site 5..110
 FT /label= His6_tag
 FT /note= "Nickel chelating region used for purification of
 FT the protein"
 FT Cleavage-site 11..116
 FT /label= Factor_X_cleavage_site
 XX US6423680-B1.
 XX 23-JUL-2002.
 XX 30-OCT-1998; 98US-00183841.
 XX 30-OCT-1998; 98US-00183841.
 XX (HSCR-) HSC RES & DEV LP.
 XX Rigat B, Reynaud D, Mahuran D;
 XX WPI; 2002-664636/71.
 XX Composition useful for treating inflammatory conditions e.g. asthma
 XX comprises GM2 activator protein or GM2 activator peptide in combination
 XX with a carrier.
 XX Example 1; Fig 2; 1lpp; English.
 XX The present invention relates to a composition comprising GM2
 CC (ganglioside) activator protein or a GM2 activator peptide derived from
 CC the GM2 activator protein in combination with a carrier. The composition
 CC comprises the protein or peptide in an amount of 1-100 mg. The GM2
 CC activator protein is capable of inhibiting platelet activating factor
 CC (PAF). The composition of the invention is useful for treating
 CC inflammatory disorders e.g. inflammatory bowel disease, asthma,
 CC autoimmune disease (such as lupus), hypersensitivity infection,
 CC rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
 CC gout and tissue-specific conditions (e.g. glomerulonephritis and
 CC hepatitis). The composition is capable of inhibiting platelet activating
 CC factor, is non-toxic, is efficacious and presents less severe side
 CC effects, including redness, swelling, pain and polymorphonuclear
 CC leukocyte accumulation at the inflammatory site and other associated
 CC cellular responses. The present sequence represents a non-glycosylated
 CC human GM2 activator protein prepared using a His6-tag bacterial
 CC expression system
 XX SQ Sequence 178 AA;
 Query Match 90.4%; Score 85; DB 5; Length 178;
 Best Local Similarity 93.8%; Pred. No. 8.5e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Best Local Similarity 93.8%; Pred. No. 7.9e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FSWDNCFEKDPKPAVIR 16
 ||||| |||||
 Db 19 FSWDNCDEGKDPKPAVIR 34
 RESULT 5
 AAB31900
 ID AAB31900 standard; protein; 189 AA.
 XX AC AAB31900;
 XX DT 15-MAY-2001 (first entry)
 XX DE Amino acid sequence of a human protein.
 XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX OS Homo sapiens.
 XX WO200105422-A2.
 XX 25-JAN-2001.
 XX 17-JUL-2000; 2000WO-FR002057.
 XX 15-JUL-1999; 99FR-00009372.
 XX (INMR) BIOMERIEUX STELHYS.
 XX Roecklin D, Kolbe H, Charles M, Marcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 XX Detecting, preventing and treating degenerative, neurological and
 XX autoimmune diseases, particularly multiple sclerosis, using specified
 XX polypeptides or related nucleic acid or ligand.
 XX Claim 1; Page 161-162; 209pp; French.
 XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX SQ Sequence 189 AA;
 Query Match 90.4%; Score 85; DB 4; Length 189;
 Best Local Similarity 93.8%; Pred. No. 8.5e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FSWDNCFEKDPKPAVIR 16
 ||||| |||||
 Db 30 FSWDNCDEGKDPKPAVIR 45

```

RESULT 6
AAB31901
ID AAB31901 standard; protein; 193 AA.
XX
AC AAB31901;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
DR Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 162-163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
XX method of the invention. The specification describes a method which uses
XX at least one polypeptide or polynucleotide sequence belonging to the
XX perlecan, precursor of the retinol-binding plasma protein, precursor of
XX the ganglioside GM2 activator, calgranulin B or saposin B protein
XX families. The method is used for detecting, preventing or treating a
XX degenerative, neurological and/or auto-immune disease. The
XX polynucleotides and polypeptides are used for diagnosis, prognosis,
XX prevention and treatment of multiple sclerosis (in its various forms and
XX phases). They may also be useful in cases of e.g. Alzheimer's and
XX Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
XX polyarthritis and lupus erythematosus, including use as vaccines and in
XX gene therapy (expression of sense or antisense sequences). They can also
XX be used to assess efficacy of potential therapeutic agents, particularly
XX compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 90.4%; Score 85; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 8.7e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFCGKDPVAVR 16
||||| |||||||
Db 34 FSWDNCDEGKDPVAVR 49

RESULT 7
AAB31902
ID AAB31902 standard; protein; 193 AA.
XX
AC AAB31902;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
DR Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 162-163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
XX method of the invention. The specification describes a method which uses
XX at least one polypeptide or polynucleotide sequence belonging to the
XX perlecan, precursor of the retinol-binding plasma protein, precursor of
XX the ganglioside GM2 activator, calgranulin B or saposin B protein
XX families. The method is used for detecting, preventing or treating a
XX degenerative, neurological and/or auto-immune disease. The
XX polynucleotides and polypeptides are used for diagnosis, prognosis,
XX prevention and treatment of multiple sclerosis (in its various forms and
XX phases). They may also be useful in cases of e.g. Alzheimer's and
XX Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
XX polyarthritis and lupus erythematosus, including use as vaccines and in
XX gene therapy (expression of sense or antisense sequences). They can also
XX be used to assess efficacy of potential therapeutic agents, particularly
XX compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 90.4%; Score 85; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 8.7e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFCGKDPVAVR 16
||||| |||||||
Db 34 FSWDNCDEGKDPVAVR 49

RESULT 8
AAB31904
ID AAB31904 standard; protein; 193 AA.
XX
AC AAB31904;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
DR Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
XX method of the invention. The specification describes a method which uses
XX at least one polypeptide or polynucleotide sequence belonging to the
XX perlecan, precursor of the retinol-binding plasma protein, precursor of
XX the ganglioside GM2 activator, calgranulin B or saposin B protein
XX families. The method is used for detecting, preventing or treating a
XX degenerative, neurological and/or auto-immune disease. The
XX polynucleotides and polypeptides are used for diagnosis, prognosis,
XX prevention and treatment of multiple sclerosis (in its various forms and
XX phases). They may also be useful in cases of e.g. Alzheimer's and
XX Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
XX polyarthritis and lupus erythematosus, including use as vaccines and in
XX gene therapy (expression of sense or antisense sequences). They can also
XX be used to assess efficacy of potential therapeutic agents, particularly
XX compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 90.4%; Score 85; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 8.7e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFCGKDPVAVR 16
||||| |||||||
Db 34 FSWDNCDEGKDPVAVR 49

RESULT 9
AAB31905
ID AAB31905 standard; protein; 193 AA.
XX
AC AAB31905;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
DR Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 162-163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
XX method of the invention. The specification describes a method which uses
XX at least one polypeptide or polynucleotide sequence belonging to the
XX perlecan, precursor of the retinol-binding plasma protein, precursor of
XX the ganglioside GM2 activator, calgranulin B or saposin B protein
XX families. The method is used for detecting, preventing or treating a
XX degenerative, neurological and/or auto-immune disease. The
XX polynucleotides and polypeptides are used for diagnosis, prognosis,
XX prevention and treatment of multiple sclerosis (in its various forms and
XX phases). They may also be useful in cases of e.g. Alzheimer's and
XX Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
XX polyarthritis and lupus erythematosus, including use as vaccines and in
XX gene therapy (expression of sense or antisense sequences). They can also
XX be used to assess efficacy of potential therapeutic agents, particularly
XX compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 90.4%; Score 85; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 8.7e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFCGKDPVAVR 16
||||| |||||||
Db 34 FSWDNCDEGKDPVAVR 49

```

OS Homo sapiens.
 XX PN WO200105422-A2.
 XX PD 25-JAN-2001.
 XX PF 17-JUL-2000; 2000WO-FR002057.
 XX PR 15-JUL-1999; 99FR-00009372.
 XX PA (INMR) BIOMERIEUX STELHYS.
 XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX DR WPI; 2001-159475/16.
 XX PT Detecting, preventing and treating degenerative, neurological and
 XX PT autoimmune diseases, particularly multiple sclerosis, using specified
 XX PT polypeptides or related nucleic acid or ligand.
 XX PS Claim 1; Page 164-165; 209pp; French.
 XX CC The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX SQ Sequence 193 AA;
 Query Match 90.4%; Score 85; DB 4; Length 193;
 Best Local Similarity 93.8%; Pred. No. 8.7e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FSWDNCFEKDPKPAVIR 16
 Db ||||| ||||| |||||
 34 FSWDNCDEGKDPKPAVIR 49
 RESULT 9
 AAB31896
 ID AAB31896 standard; protein; 193 AA.
 AC AAB31896;
 XX 15-MAY-2001 (first entry)
 XX DE Amino acid sequence of the human ganglioside GM2 activator protein.
 XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX OS Homo sapiens.
 XX PN WO200105422-A2.
 XX PD 25-JAN-2001.
 XX PF 17-JUL-2000; 2000WO-FR002057.

PR 15-JUL-1999; 99FR-00009372.
 XX (INMR) BIOMERIEUX STELHYS.
 XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX DR WPI; 2001-159475/16.
 XX DR N-PSDB; AAF54698.
 XX PT Detecting, preventing and treating degenerative, neurological and
 XX PT autoimmune diseases, particularly multiple sclerosis, using specified
 XX PT polypeptides or related nucleic acid or ligand.
 XX PS Claim 23; Page 158-159; 209pp; French.
 XX CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX SQ Sequence 193 AA;
 Query Match 90.4%; Score 85; DB 4; Length 193;
 Best Local Similarity 93.8%; Pred. No. 8.7e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FSWDNCFEKDPKPAVIR 16
 Db ||||| ||||| |||||
 34 FSWDNCDEGKDPKPAVIR 49
 RESULT 10
 AAB31928
 ID AAB31928 standard; protein; 193 AA.
 XX AC AAB31928;
 XX 15-MAY-2001 (first entry)
 XX DE Amino acid sequence of the human ganglioside GM2 activator protein.
 XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX OS Homo sapiens.
 XX PN WO200105422-A2.
 XX PD 25-JAN-2001.
 XX PF 17-JUL-2000; 2000WO-FR002057.
 XX PR 15-JUL-1999; 99FR-00009372.
 XX PA (INMR) BIOMERIEUX STELHYS.
 XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX DR WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Disclosure; Fig 1; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 90.4%; Score 85; DB 4; Length 193;
 Best Local Similarity 93.8%; Pred. No. 8.7e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGDPAVIR 16
 ||||| |||||
 Db 34 FSWDNCDEKGDPAVIR 49

RESULT 11
 AAB31903
 ID AAB31903 standard; protein; 193 AA.
 XX
 AC AAB31903;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of a human protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 DR WPI; 2001-159475/16.
 XX
 CC Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 XX Claim 1; Page 164; 209pp; French.
 XX
 CC The present sequence represents a human protein, which is used in the

CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 193 AA;

Query Match 90.4%; Score 85; DB 4; Length 193;
 Best Local Similarity 93.8%; Pred. No. 8.7e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGDPAVIR 16
 ||||| |||||
 Db 34 FSWDNCDEKGDPAVIR 49

RESULT 12
 ABG00720
 ID ABG00720 standard; protein; 193 AA.
 XX
 AC ABG00720;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #711.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS64907.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 31079; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food

CC patient, for manufacture of a medicament for treating hypoxia-regulated
 CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
 CC biological response to hypoxia conditions, or hypoxic-associated
 CC pathology in a patient. The coding sequences and proteins are also useful
 CC for monitoring the therapeutic treatment of a disease or physiological
 CC condition, such as cancer, ischaemic conditions, reperfusion injury,
 CC retinopathy, neonatal stress, preclampsia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss
 XX
 SQ Sequence 193 AA;

Query Match 90.4%; Score 85; DB 5; Length 193;
 Best Local Similarity 93.8%; Pred. No. 8.7e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVR 16
 ||||| ||||| |||||
 Db 34 FSWDNCDEGKDPVAVR 49

RESULT 15

ADN95858
 ID ADN95858 standard; protein; 193 AA.

XX AC ADN95858;

XX DT 01-JUL-2004 (first entry)

XX DE Human BRC/LEC-related protein sequence SeqID782.

XX KW growth; differentiation; blood endothelial cell; BRC;
 KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
 KW lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;
 KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
 KW inflammatory disease; cancer metastasis; lymphatic system; human.

XX OS Homo sapiens.

XX PN WO2003080640-A1.

XX PD 02-OCT-2003.

XX PF 07-MAR-2003; 2003WO-US006900.

XX PR 07-MAR-2002; 2002US-0363019P.

XX PA (LUDW-) LUDWIG INST CANCER RES.
 XX PA (LICN) LICENTIA LTD.

XX PI Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;

XX DR WPI; 2003-876899/81.

XX DR N-PSDB; ADN95859.

XX PS Example 1; SEQ ID NO 782; 176pp; English.

XX CC This invention relates to a method of differentially modulating the
 CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
 CC endothelial cells (LEC) comprising contacting endothelial cells with a
 CC composition comprising an agent that differentially modulates blood or
 CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
 CC identifying a human subject with lymphoedema and with a mutation in at
 CC least one allele of a gene encoding a LEC protein, where the mutation
 CC correlates with lymphoedema in human subjects, and with the proviso that
 CC the LEC protein is not VEGFR-3; and administering to the subject a
 CC composition comprising a lymphatic growth agent selected from VEGF-C or
 CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
 CC the development of compounds with an antiangiogenic, cytostatic,
 CC vasotropic or antiinflammatory activity or for gene therapy. The method
 CC is useful in modulating the growth or differentiation of blood
 CC endothelial cells or lymphatic endothelial cells, in treating hereditary
 CC lymphoedema, in screening for an endothelial cell disorder or
 CC predisposition to the disorder or in monitoring the efficacy or toxicity

CC of a drug on endothelial cells. The agent is useful in manufacturing a
 CC medicament for the differential modulation of blood vessel endothelial
 CC cell or lymphatic vessel endothelial cell growth or differentiation. The
 CC lymphatic growth agent may also be used in manufacturing a medicament for
 CC the treatment of hereditary lymphoedema resulting from a mutation in a
 CC LEC gene or of other diseases involving the lymphatic vessels, such as
 CC various inflammatory diseases and cancer metastasis via the lymphatic
 CC system. The present sequence is that of a human LEC/BEC differentially
 CC expressed protein which is related to the method of the invention. Note:
 CC This sequence does not appear in the specification but was obtained by
 CC the indexer using the source data given in table 14 of the specification.

XX SQ Sequence 193 AA;

Query Match 90.4%; Score 85; DB 7; Length 193;
 Best Local Similarity 93.8%; Pred. No. 8.7e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVR 16

||||| ||||| |||||
 Db 34 FSWDNCDEGKDPVAVR 49

Search completed: November 17, 2004, 07:10:01

Job time : 6.65196 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 06:43:37 ; Search time 1.13357 Seconds
(without alignments)
936.062 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 94

Sequence: 1 FSWDNCFEKDPVIR 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pap:*
- 2: /cgn2_6/ptodata/1/iaa/5B COMB.pap:*
- 3: /cgn2_6/ptodata/1/iaa/6A COMB.pap:*
- 4: /cgn2_6/ptodata/1/iaa/6B COMB.pap:*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pap:*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	90.4	178	4	US-09-183-841-2
2	85	90.4	193	4	US-09-183-841-1
3	44	46.8	406	4	US-09-599-360B-98
4	44	46.8	406	4	US-10-140-002-66
5	44	46.8	406	4	US-10-140-002-258
6	44	46.8	1765	4	US-09-270-767-45587
7	43	45.7	15	4	US-09-069-827A-129
8	43	45.7	590	2	US-08-756-317-12
9	43	45.7	619	4	US-09-489-039A-12704
10	42	44.7	119	4	US-09-270-767-47348
11	41	43.6	113	4	US-10-158-847-113
12	41	43.6	506	2	US-08-635-066-2
13	41	43.6	864	4	US-10-101-464A-896
14	41	43.6	956	4	US-09-252-991A-17124
15	40	42.6	19	1	US-08-469-615-18
16	40	42.6	19	1	US-08-466-763-18
17	40	42.6	19	2	US-08-411-142A-18
18	40	42.6	40	1	US-07-901-874B-5
19	40	42.6	40	1	US-08-457-865-5
20	40	42.6	304	4	US-09-710-279-2190
21	40	42.6	376	4	US-09-270-767-33947
22	40	42.6	376	4	US-09-270-767-49164
23	40	42.6	385	2	US-08-516-801-2
24	40	42.6	385	3	US-08-248-355-2
25	40	42.6	385	4	US-09-167-206-16
26	40	42.6	385	5	PCT-US95-06683-2
27	40	42.6	415	3	US-09-114-001C-5077

28	40	42.6	488	1	US-07-672-483-1	Sequence 1, Appli
29	40	42.6	490	2	US-08-687-916-24	Sequence 24, Appli
30	40	42.6	490	3	US-09-138-614-24	Sequence 24, Appli
31	40	42.6	619	4	US-09-543-681A-5503	Sequence 5503, Ap
32	40	42.6	621	2	US-08-969-714-1	Sequence 1, Appli
33	40	42.6	645	2	US-08-969-714-3	Sequence 3, Appli
34	39	41.5	13	4	US-10-158-847-105	Sequence 105, App
35	39	41.5	191	4	US-09-489-039A-12833	Sequence 12833, A
36	39	41.5	242	4	US-09-270-767-32046	Sequence 32046, A
37	39	41.5	364	3	US-09-338-671-2	Sequence 2, Appli
38	39	41.5	444	1	US-08-483-140-28	Sequence 28, Appli
39	39	41.5	444	2	US-08-485-938A-32	Sequence 32, Appli
40	39	41.5	658	4	US-09-328-599A-2	Sequence 2, Appli
41	39	41.5	795	3	US-09-031-563-23	Sequence 23, Appli
42	39	41.5	795	4	US-09-392-277-23	Sequence 23, Appli
43	39	41.5	795	4	US-09-258-000-23	Sequence 23, Appli
44	39	41.5	878	4	US-09-556-706B-2	Sequence 2, Appli
45	39	41.5	878	4	US-09-724-418A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-183-841-2
; Sequence 2, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: His tag at residues 1 to 17
; OTHER INFORMATION: Description of Artificial Sequence: amino acid
; OTHER INFORMATION: sequence of GM2 protein using His6 tag
US-09-183-841-2

Query Match 90.4%; Score 85; DB 4; Length 178;
Best Local Similarity 93.8%; Pred. No. 2.9e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FSWDNCFEKDPVIR 16
| | | | | | | | | | | | | | | | | |
Db 19 FSWDNCDEKDPVIR 34

RESULT 2
US-09-183-841-1
; Sequence 1, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL

```

; LOCATION: (33)...(55)
; FEATURE:
; OTHER INFORMATION: residues 56-63 are included in a further precursor
; OTHER INFORMATION: form of the protein
US-09-183-841-1
Query Match          90.4%; Score 85; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 3;le-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVR 16
    ||||| ||||| |||||
Db 34 FSWDNCDEGKDPVAVR 49

RESULT 3
US-09-599-360B-98
; Sequence 98, Application US/09599360B
; Patent No. 6548633
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Bougueleret, L.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal Peptides
; FILE REFERENCE: GENSET.050CP3
; CURRENT APPLICATION NUMBER: US/09/599,360B
; CURRENT FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: 60/113,686
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/141,032
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 09/469,099
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: Patent.pm
; SEQ ID NO 98
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -35...-1
US-09-599-360B-98

Query Match          46.8%; Score 44; DB 4; Length 406;
Best Local Similarity 46.7%; Pred. No. 32;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVI 15
    ||||| :||| :
Db 155 FSWNITDSLDPATL 169

RESULT 4
US-10-140-002-66
; Sequence 66, Application US/10140002
; Patent No. 6725730
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
US-10-140-002-66

Query Match          46.8%; Score 44; DB 4; Length 406;
Best Local Similarity 46.7%; Pred. No. 32;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVI 15
    ||||| :||| :
Db 155 FSWNITDSLDPATL 169

RESULT 5
US-10-140-002-258
; Sequence 258, Application US/10140002
; Patent No. 6725730
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC59
; CURRENT APPLICATION NUMBER: US/10/140,002
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 258
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-002-258

Query Match          46.8%; Score 44; DB 4; Length 406;
Best Local Similarity 46.7%; Pred. No. 32;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVI 15
    ||||| :||| :
Db 155 FSWNITDSLDPATL 169

RESULT 6
US-09-270-767-45587
; Sequence 45587, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:

```

```

; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC59
; CURRENT APPLICATION NUMBER: US/10/140,002
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 66
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-002-66

Query Match          46.8%; Score 44; DB 4; Length 406;
Best Local Similarity 46.7%; Pred. No. 32;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVI 15
    ||||| :||| :
Db 155 FSWNITDSLDPATL 169

RESULT 5
US-10-140-002-258
; Sequence 258, Application US/10140002
; Patent No. 6725730
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC59
; CURRENT APPLICATION NUMBER: US/10/140,002
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 258
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-002-258

Query Match          46.8%; Score 44; DB 4; Length 406;
Best Local Similarity 46.7%; Pred. No. 32;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVI 15
    ||||| :||| :
Db 155 FSWNITDSLDPATL 169

RESULT 6
US-09-270-767-45587
; Sequence 45587, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:

```


; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270.767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 45587
 ; LENGTH: 1765
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; FEATURE:
 ; OTHER INFORMATION: Xaa means any amino acid
 US-09-270-767-45587

Query Match 46.8%; Score 44; DB 4; Length 1765;
 Best Local Similarity 70.0%; Pred. No. 1.5e+02;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WDNCFEGKDP 12
 Db 718 WSGFFEGKDP 727

RESULT 7
 US-09-069-827A-129
 ; Sequence 129, Application US/09069827A
 ; Patent No. 6617114
 ; GENERAL INFORMATION:
 ; APPLICANT: FOWLKES, Dana M
 ; PRELINGER, Jeffrey A
 ; HYDE-DREYUSCHER, Robin P
 ; TITLE OF INVENTION: IDENTIFICATION OF DRUGS USING
 ; COMPLEMENTARY COMBINATORIAL LIBRARIES
 ; NUMBER OF SEQUENCES: 178
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
 ; STREET: 624 Ninth Street N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20001
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/069,827A
 ; FILING DATE: 30-Apr-1998
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/050,359
 ; FILING DATE: 31-MAR-1998
 ; APPLICATION NUMBER: PCT/US97/19638
 ; FILING DATE: 31-OCT-1997
 ; APPLICATION NUMBER: US 08/740,671
 ; FILING DATE: 31-OCT-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: COOPER, Iver P
 ; REGISTRATION NUMBER: 28,005
 ; REFERENCE/DOCKET NUMBER: FOWLKES-4C
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 129:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 15 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide

; SEQUENCE DESCRIPTION: SEQ ID NO: 129:
 US-09-069-827A-129
 Query Match 45.7%; Score 43; DB 4; Length 15;
 Best Local Similarity 60.0%; Pred. No. 1.4;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 FSWDNCFEGK 10
 Db 6 FFWDGCFESR 15
 RESULT 8
 US-08-756-317-12
 ; Sequence 12, Application US/08756317
 ; Patent No. 5849894
 ; GENERAL INFORMATION:
 ; APPLICANT: Clemente, Thomas E.
 ; APPLICANT: Kishore, Ganesh M.
 ; APPLICANT: Mitsky, Timothy A.
 ; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
 ; TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthase
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: TX
 ; COUNTRY: USA
 ; ZIP: 77210-4433
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/756,317
 ; FILING DATE: 25-NOV-1996
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/007,693
 ; FILING DATE: 29-NOV-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Patterson, Melinda L.
 ; REGISTRATION NUMBER: 33,062
 ; REFERENCE/DOCKET NUMBER: MOBT:008
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (713) 787-1400
 ; TELEFAX: (713) 787-1440
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 590 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-08-756-317-12

Query Match 45.7%; Score 43; DB 2; Length 590;
 Best Local Similarity 66.7%; Pred. No. 69;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WDNCFEGKD 11
 Db 486 WDTCFEGAD 494

RESULT 9
 US-09-489-039A-12704
 ; Sequence 12704, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

;; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

;; FILE REFERENCE: 2709.2004001

;; CURRENT APPLICATION NUMBER: US/09/489,039A

;; CURRENT FILING DATE: 2000-01-27

;; PRIOR APPLICATION NUMBER: US 60/117,747

;; PRIOR FILING DATE: 1999-01-29

;; NUMBER OF SEQ ID NOS: 14342

;; SEQ ID NO 12704

;; LENGTH: 619

;; TYPE: PRT

;; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-12704

Query Match 45.7%; Score 43; DB 4; Length 619;

Best Local Similarity 50.0%; Pred. No. 73;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 WDNCFEGKDPVIR 16

Db 478 WNNTRSGEDPAIR 491

RESULT 10

US-09-270-767-47348

;; Sequence 47348, Application US/09270767

;; Patent No. 6703491

;; GENERAL INFORMATION:

;; APPLICANT: Homburger et al.

;; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

;; FILE REFERENCE: File Reference: 7326-094

;; CURRENT APPLICATION NUMBER: US/09/270,767

;; CURRENT FILING DATE: 1999-03-17

;; NUMBER OF SEQ ID NOS: 62517

;; SOFTWARE: Patent In Ver. 2.0

;; SEQ ID NO 47348

;; LENGTH: 119

;; TYPE: PRT

;; ORGANISM: Drosophila melanogaster

US-09-270-767-47348

Query Match 44.7%; Score 42; DB 4; Length 119;

Best Local Similarity 50.0%; Pred. No. 18;

Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 WDNCFEGKDPVIR 14

Db 97 WRECFFGQPTI 108

RESULT 11

US-10-158-847-113

;; Sequence 113, Application US/10158847

;; Patent No. 6592865

;; GENERAL INFORMATION:

;; APPLICANT: Tom Parry et al.

;; TITLE OF INVENTION: Method and Compositions for Modulating ACE-2 Activity

;; FILE REFERENCE: PF557

;; CURRENT APPLICATION NUMBER: US/10/158,847

;; CURRENT FILING DATE: 2002-06-03

;; PRIOR APPLICATION NUMBER: 60/295,004

;; PRIOR FILING DATE: 2001-06-04

;; NUMBER OF SEQ ID NOS: 158

;; SOFTWARE: Patent In version 3.1

;; SEQ ID NO 113

;; LENGTH: 13

;; TYPE: PRT

;; ORGANISM: homo sapiens

US-10-158-847-113

Query Match 43.6%; Score 41; DB 4; Length 13;

Best Local Similarity 66.7%; Pred. No. 2.5;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 FSWDNCFEG 9

Db 5 FWDDECFGL 13

RESULT 12

US-08-635-066-2

;; Sequence 2, Application US/08635066

;; Patent No. 5945580

;; GENERAL INFORMATION:

;; APPLICANT: Dunsuir, Pamela

;; APPLICANT: Harpster, Mark H.

;; TITLE OF INVENTION: Capsicum Hemicellulase Polynucleotides

;; TITLE OF INVENTION: and Polypeptides

;; NUMBER OF SEQUENCES: 8

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Townsend and Townsend and Crew LLP

;; STREET: Two Embarcadero Center, Eighth Floor

;; CITY: San Francisco

;; STATE: California

;; COUNTRY: USA

;; ZIP: 94111-3834

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patent In Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/635,066

;; FILING DATE: 19-APR-1996

;; CLASSIFICATION: 800

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Smith, William M.

;; REGISTRATION NUMBER: 30,223

;; REFERENCE/DOCKET NUMBER: 012176-005500

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (415) 576-0200

;; TELEFAX: (415) 576-0300

;; INFORMATION FOR SEQ ID NO: 2:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 506 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

US-08-635-066-2

Query Match 43.6%; Score 41; DB 2; Length 506;

Best Local Similarity 77.8%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FSWDNCFEG 9

Db 266 FSWDNCFAG 274

RESULT 13

US-10-101-464A-896

;; Sequence 896, Application US/10101464A

;; Patent No. 6768041

;; GENERAL INFORMATION:

;; APPLICANT: Strabala, Timothy

;; APPLICANT: Nieuwenhuizen, Nicolaas

;; APPLICANT: Higgins, Colleen M.

;; TITLE OF INVENTION: Compositions Isolated from Plant Cells

;; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling

;; FILE REFERENCE: 11000.1020G2

;; CURRENT APPLICATION NUMBER: US/10/101,464A

;; CURRENT FILING DATE: 2002-03-18

;; PRIOR APPLICATION NUMBER: 09/704,302

;; PRIOR FILING DATE: 2000-11-01

;; PRIOR APPLICATION NUMBER: 09/228,986

;; PRIOR FILING DATE: 1999-01-12

```
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 896
; LENGTH: 864
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-896

Query Match          43.6%; Score 41; DB 4; Length 864;
Best Local Similarity 58.3%; Pred. No. 2.2e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      5 NCFEGKDPVIR 16
Db      402 NCFSGAIPSLIR 413

RESULT 14
US-09-252-991A-17124
; Sequence 17124, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17124
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17124

Query Match          43.6%; Score 41; DB 4; Length 956;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      4 DNCFEKDP 12
Db      291 EHCFCGHDP 299

RESULT 15
US-08-469-615-18
; Sequence 18, Application US/08469615
; Patent No. 5622703
; GENERAL INFORMATION:
; APPLICANT: Berzofsky, Jay A.
; APPLICANT: Kurata, Akihiko
; TITLE OF INVENTION: Immunodominant Sites of HTLV-I Envelope
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,615
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/243,118
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30330
; REFERENCE/DOCKET NUMBER: 1173-500P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: htlv-I
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..19
; OTHER INFORMATION: /label=peptid_vie9
US-08-469-615-18
```

```
Query Match          42.6%; Score 40; DB 1; Length 19;
Best Local Similarity 33.3%; Pred. No. 5.5;
Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY      1 FSWDNCFEKDPVIR 15
Db      4 FNTWTHCFDPQIAIV 18
```

```
Search completed: November 17, 2004, 07:27:57
Job time : 2.22447 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 07:01:42 : Search time 3.82191 Seconds
(without alignments)
1482.515 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 94

Sequence: 1 FSWDNCFEKDPVAVIR 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100%

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	90.4	193	14	US-10-170-385-389
2	57	60.6	126	10	Sequence 389, App
3	57	60.6	191	15	Sequence 4977, Ap
4	47	50.0	620	14	Sequence 2611, Ap
5	47	50.0	623	14	Sequence 14781, A
6	47	50.0	649	14	Sequence 14935, A
7	46	48.9	101	17	Sequence 11373, A
8	46	48.9	211	14	Sequence 246626, A
9	46	48.9	211	15	Sequence 96, Appl
10	46	48.9	588	15	Sequence 2846, Ap
11	45	47.9	493	14	Sequence 77389, A
12	45	47.9	493	14	Sequence 4327, Ap
13	45	47.9	1411	15	Sequence 7083, Ap
					Sequence 43060, A

14	44	46.8	74	16	US-10-437-963-110611	Sequence 110611,
15	44	46.8	90	17	US-10-425-115-350643	Sequence 350643,
16	44	46.8	108	17	US-10-425-115-351913	Sequence 351913,
17	44	46.8	132	9	US-09-739-907-99	Sequence 99, Appl
18	44	46.8	132	11	US-09-938-671-99	Sequence 99, Appl
19	44	46.8	172	9	US-09-739-907-87	Sequence 87, Appl
20	44	46.8	172	11	US-09-938-671-87	Sequence 87, Appl
21	44	46.8	177	11	US-09-833-245-1188	Sequence 1188, Ap
22	44	46.8	182	9	US-09-739-907-191	Sequence 191, App
23	44	46.8	182	11	US-09-938-671-191	Sequence 191, App
24	44	46.8	213	17	US-10-425-115-224812	Sequence 224812,
25	44	46.8	330	11	US-09-833-245-1189	Sequence 1189, Ap
26	44	46.8	406	9	US-09-731-872-245	Sequence 245, App
27	44	46.8	406	10	US-09-876-997-245	Sequence 245, App
28	44	46.8	406	11	US-09-978-360A-430	Sequence 430, App
29	44	46.8	406	14	US-10-028-072-66	Sequence 66, Appl
30	44	46.8	406	14	US-10-028-072-258	Sequence 258, App
31	44	46.8	406	14	US-10-140-808-66	Sequence 66, Appl
32	44	46.8	406	14	US-10-140-808-258	Sequence 258, App
33	44	46.8	406	14	US-10-121-049-66	Sequence 66, Appl
34	44	46.8	406	14	US-10-121-049-258	Sequence 258, App
35	44	46.8	406	14	US-10-123-904-66	Sequence 66, Appl
36	44	46.8	406	14	US-10-123-904-258	Sequence 258, App
37	44	46.8	406	14	US-10-140-470-66	Sequence 66, Appl
38	44	46.8	406	14	US-10-140-470-258	Sequence 258, App
39	44	46.8	406	14	US-10-175-746-66	Sequence 66, Appl
40	44	46.8	406	14	US-10-175-746-258	Sequence 258, App
41	44	46.8	406	14	US-10-176-918-66	Sequence 66, Appl
42	44	46.8	406	14	US-10-176-918-258	Sequence 258, App
43	44	46.8	406	14	US-10-176-921-66	Sequence 66, Appl
44	44	46.8	406	14	US-10-176-921-258	Sequence 258, App
45	44	46.8	406	14	US-10-137-865-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1

US-10-170-385-389
; Sequence 389, Application US/10170385
; Publication No. US2003020372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingeman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 532682000100
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 389
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-170-385-389

Query Match 90.4%; Score 85; DB 14; Length 193;
Best Local Similarity 93.8%; Pred. No. 4.2e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEKDPVAVIR 16

```

; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (187)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2611

Query Match          60.6%; Score 57; DB 15; Length 191;
Best Local Similarity 50.0%; Pred. No. 0.22;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVR 16
    | : | | | | | | | | | | | | | | | |
Db 52 FFWENCHERKDPVLK 67

RESULT 4
US-10-369-493-14781
; Sequence 14781, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14781
; LENGTH: 620
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(620)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-14781

Query Match          50.0%; Score 47; DB 14; Length 620;
Best Local Similarity 56.2%; Pred. No. 36;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 SMDNCFEG--KDPAVI 15
    | : | | | | | | | | | | | |
Db 459 SMNNGFGLTDDPAIV 474

RESULT 5
US-10-369-493-14935
; Sequence 14935, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14935
; LENGTH: 623

```

```

; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (184)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2611

Query Match          60.6%; Score 57; DB 10; Length 126;
Best Local Similarity 50.0%; Pred. No. 0.14;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVR 16
    | : | | | | | | | | | | | |
Db 52 FFWENCHERKDPVLK 67

RESULT 3
US-10-264-049-2611
; Sequence 2611, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2611
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (141)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (142)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (184)

```

```
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(623)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-14935

Query Match      50.0%; Score 47; DB 14; Length 623;
Best Local Similarity 56.2%; Pred. No. 37;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY      2 SWDNCFEK--KDPVAVI 15
Db      458 SWNNGFEGLTDDPAIV 473

RESULT 6
US-10-369-493-11373
; Sequence 11373, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11373
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11373

Query Match      50.0%; Score 47; DB 14; Length 649;
Best Local Similarity 56.2%; Pred. No. 38;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY      2 SWDNCFEK--KDPVAVI 15
Db      465 SWNNGFEGLTDDPAIV 480

RESULT 7
US-10-425-115-246626
; Sequence 246626, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 246626
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_156506C.1.pep
US-10-425-115-246626
```

```
Query Match      48.9%; Score 46; DB 17; Length 101;
Best Local Similarity 46.2%; Pred. No. 8;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY      3 WDNCFEKGKDPVAVI 15
Db      28 WNSCFHGPEPPVM 40

RESULT 8
US-10-225-066A-96
; Sequence 96, Application US/10225066A
; Publication No. US20030226173A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: RATCLIFFE, Oliver
; APPLICANT: RIECHMANN, Jose Luis
; APPLICANT: ADAM, Luc J
; APPLICANT: DUBELL, Arnold T
; APPLICANT: HEARD, Jacqueline E
; APPLICANT: PILGRIM, Marsha L
; APPLICANT: JIANG, Cai-Zhong
; APPLICANT: REUBER, T. Lynne
; APPLICANT: CREELMAN, Robert A
; APPLICANT: PINEDA, Omaira
; APPLICANT: YU, Guo-Liang
; APPLICANT: BROUN, Pierre E
; TITLE OF INVENTION: Field-Related Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MB10036-2 US
; CURRENT APPLICATION NUMBER: US/10/225,066A
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 09/837,444
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 1122
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 96
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-225-066A-96

Query Match      48.9%; Score 46; DB 14; Length 211;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      2 SWDNCFEKGD 11
Db      169 SWYNCFDGDD 178

RESULT 9
US-10-374-780A-2846
; Sequence 2846, Application US/10374780A
; Publication No. US20040019927A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
```

```
; APPLICANT: Broun, Pierre E
; APPLICANT: Pilgrim, Marsha L
; APPLICANT: Dubell III, Arnold T
; APPLICANT: Pinella, Omaira
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 CIP
; CURRENT APPLICATION NUMBER: US/10/374,780A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 09/837,944
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/934,455
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 10/225,066
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,067
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,068
; PRIOR FILING DATE: 2002-08-09
; NUMBER OF SEQ ID NOS: 2906
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2846
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G2294
US-10-374-780A-2846

Query Match 48.9%; Score 46; DB 15; Length 211;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 SWDNCPEGKD 11
Db 169 SWYCNCFDGD 178

RESULT 10
US-10-282-122A-77389
; Sequence 77389, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
```

```
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77389
; LENGTH: 588
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-282-122A-77389

Query Match 48.9%; Score 46; DB 15; Length 588;
Best Local Similarity 57.1%; Pred. No. 51;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 WDNCFEGKDPVAVIR 16
Db 447 WENXKGEDPVVIR 460

RESULT 11
US-10-369-493-4327
; Sequence 4327, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4327
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-369-493-4327

Query Match 47.9%; Score 45; DB 14; Length 493;
Best Local Similarity 58.3%; Pred. No. 62;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 WDNCFEGKDPVAV 14
Db 449 WYNCYQVMDEAV 460

RESULT 12
US-10-369-493-7083
; Sequence 7083, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
```


APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 7083
LENGTH: 493
TYPE: PRT
ORGANISM: Burkholderia cepacia
US-10-369-493-7083

Query Match 47.9%; Score 45; DB 14; Length 493;
Best Local Similarity 58.3%; Pred. No. 62;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 WDNCFEGKDPVAV 14
|||:|
Db 449 WNCYQVNDPAV 460

RESULT 13
US-10-282-122A-43060
Sequence 43060, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 43060
LENGTH: 1411
TYPE: PRT
ORGANISM: Escherichia coli

US-10-282-122A-43060
Query Match 47.9%; Score 45; DB 15; Length 1411;
Best Local Similarity 42.9%; Pred. No. 1.9e+02;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 WDNCFEGKDPVAV 16
|||:|
Db 1354 WDEIFSGNDEKIVK 1367

RESULT 14
US-10-437-963-110611
Sequence 110611, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 110611
LENGTH: 74
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_14661C.1.pep
US-10-437-963-110611

Query Match 46.8%; Score 44; DB 16; Length 74;
Best Local Similarity 33.3%; Pred. No. 12;
Matches 4; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEGKDP 12
:|:|:|:|
Db 20 YAWDSCYQQQEP 31

RESULT 15
US-10-425-115-350643
Sequence 350643, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 350643
LENGTH: 90
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_82957C.1.pep
US-10-425-115-350643

Query Match 46.8%; Score 44; DB 17; Length 90;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SWDNCFEKQPA 13
: : : : :
Db 16 AWSSGSDGPDPA 27

Search completed: November 17, 2004, 07:38:22
Job time : 4.82191 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 06:36:49 ; Search time 0.81146 Seconds
(without alignments)
1897.160 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 94

Sequence: 1 FSWDNCFEKDPVIR 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	90.4	162	2 S13195	ganglioside M2 act
2	85	90.4	193	2 I54178	ganglioside M2 act
3	85	90.4	200	2 S22411	ganglioside M2 act
4	82	87.2	193	2 S35613	ganglioside M2 act
5	60	63.8	20	2 S56005	lysosomal protein
6	47	50.0	171	2 G72563	hypothetical prote
7	47	50.0	651	2 AD3057	glycogen debranchi
8	47	50.0	651	2 B98229	hypothetical prote
9	46	48.9	211	2 D96507	glycogen debranchi
10	46	48.9	588	2 G82118	hypothetical prote
11	45	47.9	1411	2 E65145	succinate dehydrog
12	43	45.7	395	2 T20724	thbB protein precu
13	43	45.7	530	2 I39685	hypothetical prote
14	42	44.7	27	2 D44908	polyhydroxyalkanol
15	42	44.7	262	2 S23241	chitinase EC 3.2.
16	42	44.7	329	2 D97230	hypothetical prote
17	42	44.7	379	2 G81712	sugar kinase, ribo
18	42	44.7	522	2 C88543	1-deoxy-D-xylulose
19	42	44.7	807	2 C85025	protein ZK643.3 [i
20	42	44.7	823	2 B87348	hypothetical prote
21	42	44.7	871	2 B71039	1,4-Beta-D-glucan
22	41	43.6	337	2 AC0377	hypothetical prote
23	41	43.6	343	2 H84607	probable mannosylt
24	41	43.6	343	2 H84607	hypothetical prote
25	41	43.6	442	2 AI3075	glutamate synthase
26	41	43.6	442	2 H98210	glutamC protein [i
27	41	43.6	479	2 T05388	cellulase EC 3.2.
28	41	43.6	488	2 T04021	cellulase EC 3.2.
29	41	43.6	506	2 T07885	cellulase EC 3.2.
				2 S57663	cellulase EC 3.2.

ALIGNMENTS

RESULT 1

S13195

ganglioside M2 activator protein - human

C/Species: Homo sapiens (man)

C/Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004

C/Accession: S13195

R/Fuerst, W.; Schubert, J.; Machleidt, W.; Meyer, H.E.; Sandhoff, K.

Eur. J. Biochem. 192, 709-714, 1990

A/Title: The complete amino-acid sequences of human ganglioside GM2 activator protein and beta-D-glucan exoh

A/Reference number: S13195; MUID:91006165; PMID:2209618

A/Accession: S13195

A/Status: preliminary

A/Molecule type: protein

A/Residues: 1-162 <FUE>

A/Cross-references: UNIPROT:P17900

Query Match 90.4%; Score 85; DB 2; Length 162;

Best Local Similarity 93.8%; Pred. No. 2.3e-06;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 FSWDNCFEKDPVIR 16

|||||

Db

3 FSWDNCDEGKDPVIR 18

|||||

RESULT 2

I54178

ganglioside M2 activator protein precursor - human

C/Species: Homo sapiens (man)

C/Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004

C/Accession: I54178; JQ1037; S05036; S22410; S17107

R/Xie, B.; Kennedy, J.L.; McInnes, B.; Auger, D.; Mahuran, D.

Genomics 14, 796-798, 1992

A/Title: Identification of a processed pseudogene related to the functional gene encoding man chromosome 5.

A/Reference number: I54178; MUID:93052421; PMID:1427911

A/Accession: I54178

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-193 <RES>

A/Cross-references: UNIPROT:P17900; GB:L01439; NID:g183358; PIDN:AAA52767.1; PID:g183359

R/Xie, B.; McInnes, B.; Neote, K.; Lamhonwah, A.M.; Mahuran, D.

Biochem. Biophys. Res. Commun. 177, 1217-1223, 1991

A/Title: Isolation and expression of a full-length cDNA encoding the human GM2 activator

A/Reference number: JQ1037; MUID:91282768; PMID:2059210

A/Accession: JQ1037

A/Molecule type: mRNA

A/Residues: 1-18, 'A', 20-193 <XIE>

A/Cross-references: GB:M76477; NID:g183356; PIDN:AAA35907.1; PID:g183357

A/Experimental source: HeLa cell

A/Note: 19-Thr and 59-Met were also found

R/Schroeder, M.; Klima, H.; Nakano, T.; Kwon, H.; Quintern, L.E.; Gaertner, S.; Suzuki, T.

FEBS Lett. 251, 197-200, 1989
A:Title: Isolation of a cDNA encoding the human G(M2) activator protein.
A:Reference number: S05036; MUID:89325664; PMID:2753159
A:Accession: S05036
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 15-193 <SCH>
A:Cross-references: GB:X16087; NID:g31852; PIDN:CAA34215.1; PID:g31853
R:Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A:Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A:Reference number: S22410; MUID:92207171; PMID:1554364
A:Accession: S22410
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 43-142, 'VST' <NAG>
A:Cross-references: EMBL:X61094; NID:g31854; PIDN:CAA43407.1; PID:g31855
C:Comment: this protein transports GM2 ganglioside from the lysosomal membrane to hexosa
C:Genetics:
A:Gene: GDB:GM2A
A:Cross-references: GDB:120000; OMIM:272750
A:Map position: 5q32-5q33
F:1-23/Domain: signal sequence #status predicted <SIG>
F:32-193/Product: GM2 ganglioside activator protein #status predicted <GM2>
F:32-193/Product: ganglioside M2 activator #status predicted <MAT>
Query Match 90.4%; Score 85; DB 2; Length 193;
Best Local Similarity 93.8%; Pred. No. 2.7e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FSWDNCFEKDPFAVIR 16
||||| |||||||
Db 34 FSWDNCDEKDPFAVIR 49
RESULT 3
S22411
ganglioside M2 activator protein (clone pGAP2) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
A:Accession: S22411; S17108
R:Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A:Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A:Reference number: S22410; MUID:92207171; PMID:1554364
A:Accession: S22411
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-200 <NAG>
A:Cross-references: UNIPROT:P17900; EMBL:X61095; NID:g31856; PIDN:CAA43408.1; PID:g31857
Query Match 90.4%; Score 85; DB 2; Length 200;
Best Local Similarity 93.8%; Pred. No. 2.8e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FSWDNCFEKDPFAVIR 16
||||| |||||||
Db 41 FSWDNCDEKDPFAVIR 56
RESULT 4
S35613
ganglioside M2 activator protein precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 07-May-1999
A:Accession: S35613
R:Belachova, G.; Stirling, J.L.; Orlacchio, A.; Beccari, T.
Biochem. J. 294, 227-230, 1993
A:Title: Cloning and sequence analysis of a cDNA clone coding for the mouse G(M2) activa
A:Reference number: S35613; MUID:93371367; PMID:7689829
A:Accession: S35613
A>Status: preliminary
A:Molecule type: mRNA

A:Residues: 1-193 <BEL>
Query Match 87.2%; Score 82; DB 2; Length 193;
Best Local Similarity 87.5%; Pred. No. 8.4e-06;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 FSWDNCFEKDPFAVIR 16
||||| |||||||
Db 34 FSWDNCDEKDPFAVIR 49
RESULT 5
S56005
lysosomal protein 22K - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 07-May-1999
A:Accession: S56005
R:Kuwana, T.; Mullock, B.M.; Luzio, J.P.
Biochem. J. 308, 937-946, 1995
A:Title: Identification of a lysosomal protein causing lipid transfer, using a fluoresce
A:Reference number: S56005; MUID:97104296; PMID:8948454
A:Accession: S56005
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-20 <XUM>
Query Match 63.8%; Score 60; DB 2; Length 20;
Best Local Similarity 75.0%; Pred. No. 0.003;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 FSWDNCFEKDPFAVIR 16
||||| |||||||
Db 1 FSWDNCDEKDPFAVIR 16
RESULT 6
G72563
hypothetical protein APE1793 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
A:Accession: G72563
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: G72563
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-171 <KAW>
A:Cross-references: UNIPROT:Q9YB02; DDBJ:AF000062; NID:g5105244; PIDN:BAA80796.1; PID:g1
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1793
Query Match 50.0%; Score 47; DB 2; Length 171;
Best Local Similarity 37.5%; Pred. No. 3.2;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
QY 1 FSWDNCFEKDPFAVIR 16
: ||| : ||| :
Db 44 WGNWSCFSGKPLVLR 59
RESULT 7
AD3057
glycogen debranching enzyme glgX [imported] - Agrobacterium tumefaciens (strain C58, Dup
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
A:Accession: AD3057
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD3057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-651 <KUR>
A:Cross-references: UNIPROT:Q8U8L7; GB:AE008689; PIDN:AAL44874.1; PID:gl7742522; GSPDB:G
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: glgX
A:Map position: linear chromosome
C:Superfamily: isoamylase type debranching enzyme

Query Match 50.0%; Score 47; DB 2; Length 651;
Best Local Similarity 56.2%; Pred. No. 12;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 SWDNCPEG--KDPVAVI 15
||:|||||
Db 465 SWNNGFGLTDDPAIV 480

RESULT 8
B98229
glycogen debranching enzyme (A291603) [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: B98229
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: B98229
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-651 <KUR>
A:Cross-references: UNIPROT:Q8U8L7; GB:AE007870; PIDN:AAK89356.1; PID:g15159204; GSPDB:G
C:Genetics:
A:Gene: AGR L1566
A:Map position: linear chromosome
C:Superfamily: isoamylase type debranching enzyme

Query Match 50.0%; Score 47; DB 2; Length 651;
Best Local Similarity 56.2%; Pred. No. 12;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 2 SWDNCPEG--KDPVAVI 15
||:|||||
Db 465 SWNNGFGLTDDPAIV 480

RESULT 9
D96507
hypothetical protein T12C22.10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D96507
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huizar, L. Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96507

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <STO>
A:Cross-references: UNIPROT:Q9LPE8; GB:AE005173; NID:g8655993; PIDN:AAF78266.1; GSPDB:GN
C:Genetics:
A:Gene: T12C22.10
A:Map position: 1

Query Match 48.9%; Score 46; DB 2; Length 211;
Best Local Similarity 70.0%; Pred. No. 5.8;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 SWDNCPEGKD 11
|||
Db 169 SWYCNCFDGD 178

RESULT 10

G82118

succinate dehydrogenase, flavoprotein chain VC2089 [imported] - Vibrio cholerae (strain 1
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: G82118
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P. 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: G82118
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-588 <HEI>
A:Cross-references: UNIPROT:Q9KQB1; GB:AE004283; GB:AE003852; NID:g9656636; PIDN:AAF95235
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2089
A:Map position: 1
C:Superfamily: fumarate reductase flavoprotein; 3-oxosteroid 1-dehydrogenase homology; fr

Query Match 48.9%; Score 46; DB 2; Length 588;
Best Local Similarity 57.1%; Pred. No. 16;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 WDNCPEGKDPVAVIR 16
||:|||||
Db 447 WENSKGDDPVWIR 460

RESULT 11

E65145

rnsB protein precursor - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004

C:Accession: E65145; S47701; B36902; A30092; I34935

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col

A.; Rose, D.J.; Mau, B.; Shaq, Y. Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: E65145

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1411 <BLAT>

A:Cross-references: UNIPROT:PI6917; UNIPROT:PI6918; GB:AE000424; GB:U00096; NID:g2367230;

A:Experimental source: strain K-12, substrain MG1655

R:Plunkett, G. submitted to the EMBL Data Library, March 1994

A:Reference number: S47666

A:Accession: S47701

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1129, 'Q', 1131-1411 <PLU>

Fri Nov 19 14:12:31 2004

A;Cross-references: EMBL:U00039; NID:9466582; PIDN:AB18457.1; PID:9466618
A;Experimental source: strain K-12, substrain MG1655
R;Zhao, S.; Sandt, C.H.; Feulner, G.; Vlazny, D.A.; Gray, J.A.; Hill, C.W.
J. Bacteriol. 175, 2799-2808, 1993
A;Title: Rhs elements of *Escherichia coli* K-12: complex composites of shared and unique
A;Reference number: A36902; MUID:93259920; PMID:8387990
A;Accession: B36902
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1397-1411 <ZHA>
A;Note: sequence extracted from NCBI backbone (NCBIN:132073, NCBI:132075)
R;Sadovsky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.
J. Bacteriol. 171, 636-642, 1989
A;Title: rhs gene family of *Escherichia coli* K-12.
A;Reference number: A91901; MUID:89123133; PMID:2644231
A;Accession: A30092
A;Molecule type: DNA
A;Residues: 1-100 <SAD>
R;Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehnert, A.F.; Sadovsky, A.B.; Vlazny, D.A.;
J. Bacteriol. 172, 446-456, 1990
A;Title: Structure of the rhsA locus from *Escherichia coli* K-12 and comparison of rhsA
A;Reference number: I54935; MUID:90094253; PMID:2403547
A;Accession: I54935
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1221-1411 <RES>
A;Cross-references: GB:M29717; NID:9147622; PID:9147623
C;Comment: the rhs core consist of two distinct parts: a large N-terminal core that is c
C;Genetics:
A;Gene: rhsB
A;Map position: 77 min
C;Superfamily: rhsF protein
C;Keywords: transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1411/Product: rhsD protein #status predicted <MAT>
F;148-55/Domain: transmembrane #status predicted <TM>

Query Match 47.9%; Score 45; DB 2; Length 1411;
Best Local Similarity 42.9%; Pred. No. 56;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 WDNCPEGKPAVIR 16
DB 1354 WDEIFSGKDEKIVK 1367

RESULT 12
T20724
hypothetical protein F10G8.4 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T20724
R;Basham, V.
submitted to the EMBL Data Library, September 1996
A;Reference number: Z19315
A;Accession: T20724
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-395 <WIL>
A;Cross-references: UNIPROT:Q93455; EMBL:Z80216; PIDN:CAB02281.1; GSPDB:GNO0019; CESP:EF
C;Genetics:
A;Gene: CESP:F10G8.4
A;Map position: 1
A;Introns: 32/3; 68/1; 176/3; 221/1; 384/3

Query Match 45.7%; Score 43; DB 2; Length 395;
Best Local Similarity 46.7%; Pred. No. 33;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 FSWDNCPEGKDPVAVI 15
DB 113 FFQWVFEQKSPVAVI 127

A;Cross-references: EMBL:U00039; NID:9466582; PIDN:AB18457.1; PID:9466618
A;Experimental source: strain K-12, substrain MG1655
R;Zhao, S.; Sandt, C.H.; Feulner, G.; Vlazny, D.A.; Gray, J.A.; Hill, C.W.
J. Bacteriol. 175, 2799-2808, 1993
A;Title: Rhs elements of *Escherichia coli* K-12: complex composites of shared and unique
A;Reference number: A36902; MUID:93259920; PMID:8387990
A;Accession: B36902
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1397-1411 <ZHA>
A;Note: sequence extracted from NCBI backbone (NCBIN:132073, NCBI:132075)
R;Sadovsky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.
J. Bacteriol. 171, 636-642, 1989
A;Title: rhs gene family of *Escherichia coli* K-12.
A;Reference number: A91901; MUID:89123133; PMID:2644231
A;Accession: A30092
A;Molecule type: DNA
A;Residues: 1-100 <SAD>
R;Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehnert, A.F.; Sadovsky, A.B.; Vlazny, D.A.;
J. Bacteriol. 172, 446-456, 1990
A;Title: Structure of the rhsA locus from *Escherichia coli* K-12 and comparison of rhsA
A;Reference number: I54935; MUID:90094253; PMID:2403547
A;Accession: I54935
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1221-1411 <RES>
A;Cross-references: GB:M29717; NID:9147622; PID:9147623
C;Comment: the rhs core consist of two distinct parts: a large N-terminal core that is c
C;Genetics:
A;Gene: rhsB
A;Map position: 77 min
C;Superfamily: rhsF protein
C;Keywords: transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1411/Product: rhsD protein #status predicted <MAT>
F;148-55/Domain: transmembrane #status predicted <TM>

RESULT 13

I39685
polyhydroxyalkanoic acid synthase - *Acinetobacter* sp.
C;Species: *Acinetobacter* sp.
C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 09-Jul-2004
C;Accession: I39685
R;Schembri, M.A.; Bayly, R.C.; Davies, J.K.
FEMS Microbiol. Lett. 118, 145-152, 1994
A;Title: Cloning and analysis of the polyhydroxyalkanoic acid synthase gene from an *Acin*
A;Reference number: I39685; MUID:94283855; PMID:8013870
A;Accession: I39685
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-590 <RES>
A;Cross-references: UNIPROT:Q57164; EMBL:U04848; NID:9443735; PIDN:AAA52191.1; PID:94437
C;Genetics:
A;Gene: phaC
C;Superfamily: poly(3-hydroxyalkanoic acid) synthase phbC

Query Match 45.7%; Score 43; DB 2; Length 590;
Best Local Similarity 66.7%; Pred. No. 49;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WDNCPEGKD 11
DB 485 WDTCFRGAD 494

RESULT 14

D44908
chitinase (EC 3.2.1.14), 70K - *Streptomyces olivaceoviridis* (fragment)
C;Species: *Streptomyces olivaceoviridis*
C;Date: 01-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 26-May-2000
C;Accession: D44908
R;Romaguera, A.; Menge, U.; Breves, R.; Diekmann, H.
J. Bacteriol. 174, 3450-3454, 1992
A;Title: Chitinases of *Streptomyces olivaceoviridis* and significance of processing for
A;Reference number: A44908; MUID:92276319; PMID:1592803
A;Accession: D44908
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-27 <ROM>
A;Experimental source: ATCC 11238
C;Superfamily: *Streptomyces plicatus* chitinase 63; bacterial cellulose-binding domain h

Query Match 44.7%; Score 42; DB 2; Length 27;
Best Local Similarity 75.0%; Pred. No. 3.2;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 WDNCPEGK 10
DB 14 WGSCEGK 21

RESULT 15

S23241
hypothetical protein ZK643.3 - *Caenorhabditis elegans* (fragment)
C;Species: *Caenorhabditis elegans*
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: S23241
R;Sulston, J.; Du, Z.; Thomas, K.; Wilson, R.; Hillier, L.; Staden, R.; Halloran, N.;
submitted to the EMBL Data Library, October 1991
A;Reference number: S23239
A;Accession: S23241
A;Molecule type: DNA
A;Residues: 1-262 <SUL>
A;Cross-references: UNIPROT:P30650; EMBL:Z11126
C;Genetics:

Query Match 44.7%; Score 42; DB 2; Length 27;
Best Local Similarity 75.0%; Pred. No. 3.2;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 WDNCPEGK 10
DB 14 WGSCEGK 21

A; Introns: 58/1; 135/2
C; Keywords: G protein-coupled receptor; transmembrane protein

Query Match 44.7%; Score 42; DB 2; Length 262;
Best Local Similarity 43.8%; Pred. No. 32;
Matches 7; Conservative 3; Mismatches 4; Indels 2; Gaps 1;

QY 2 SWD--NCFEGKDPAYI 15
:|||:|:
Db 21 TWDGWNCFDSATPGVV 36

Search completed: November 17, 2004, 07:24:54
Job time : 3.81146 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 06:33:32 ; Search time 4.63957 Seconds
(without alignments)
1984.236 Million cell updates/sec

Title: US-10-030-937-68

Perfect score: 94

Sequence: 1 FSWDNCPEGKDPVIR 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt 02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	90.4	47	Q8I028	Q8I028 macaca fasc
2	85	90.4	189	Q6LBI5	Q6LBI5 homo sapien
3	85	90.4	189	CAA43994	CAA43994 homo sapi
4	85	90.4	190	Q8HXX6	Q8HXX6 macaca fasc
5	85	90.4	193	1 SAP3 HUMAN	P17900 homo sapien
6	82	87.2	193	1 SAP3 MOUSE	Q60648 mus musculu
7	82	87.2	199	Q6IN37	Q6IN37 rattus norv
8	82	87.2	199	Q8CJH4	Q8CJH4 rattus norv
9	60	63.8	20	Q9QUW2	Q9QUW2 rattus sp.
10	52	55.3	338	Q8CDW5	Q8CDW5 m mus muscu
11	52	55.3	347	Q8CCH2	Q8CCH2 m mus muscu
12	48	51.1	565	Q89D38	Q89D38 bradyrhizob
13	48	51.1	2802	Q6BYI9	Q6BYI9 debaryomyce
14	47	50.0	171	Q9YB02	Q9YB02 aeropyrum p
15	47	50.0	282	Q7T362	Q7T362 brachydanio
16	47	50.0	651	Q7CU05	Q7CU05 agrobacteri
17	47	50.0	651	Q8U8L7	Q8U8L7 agrobacteri
18	46	48.9	211	Q9LPE8	Q9LPE8 arabidopsis
19	46	48.9	490	Q7XB15	Q7XB15 lilium long
20	46	48.9	580	Q91LH9	Q91LH9 white spot
21	46	48.9	581	Q8V897	Q8V897 white spot
22	46	48.9	588	Q9KQB1	Q9KQB1 vibrio chol
23	45	47.9	137	Q855S2	Q855S2 mycobacteri
24	45	47.9	145	Q7YCW4	Q7YCW4 arion lusit
25	45	47.9	145	Q7YCW5	Q7YCW5 arion lusit
26	45	47.9	1411	1 RHB5 ECOLI	P16917 escherichia
27	45	47.9	1411	2 Q46748	Q46748 escherichia
28	45	47.9	1612	2 Q83DF0	Q83DF0 coxiella bu
29	44	46.8	305	2 Q8CCW6	Q8CCW6 mus musculu
30	44	46.8	392	2 Q8IV07	Q8IV07 homo sapien
31	44	46.8	406	2 Q6UX96	Q6UX96 homo sapien

32	44	46.8	406	2	Q8WNB7	Q8WNB7 homo sapien
33	44	46.8	406	2	AAQ88815	AAQ88815 homo sapi
34	44	46.8	415	2	Q96F65	Q96F65 homo sapien
35	44	46.8	490	2	Q8FP18	Q8FP18 corynebacte
36	44	46.8	561	2	Q75E93	Q75E93 ashbya goss
37	44	46.8	561	2	AA50554	AA50554 ashbya go
38	44	46.8	747	2	Q94GH4	Q94GH4 oryza sativ
39	44	46.8	758	2	Q8IGI6	Q8IGI6 oryza sativ
40	44	46.8	778	2	Q8IQP9	Q8IQP9 drosophila
41	44	46.8	1008	2	Q9VVC5	Q9VVC5 drosophila
42	44	46.8	1013	2	Q7PIB2	Q7PIB2 anopheles g
43	44	46.8	1017	2	Q8IQ00	Q8IQ00 drosophila
44	44	46.8	1017	2	AA93737	AA93737 drosophila
45	44	46.8	1018	2	Q7PIB4	Q7PIB4 anopheles g

ALIGNMENTS

RESULT 1				
Q8I028				
ID	Q8I028	PRELIMINARY;	PRT;	47 AA.
AC	Q8I028;			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	Ganglioside GM2 activator (fragment).			
GN	Name=gm2a;			
OS	Macaca fascicularis (Crab eating macaque) (Cynomolgus macaque)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;			
OC	Cercopitheidae; Macaca.			
OX	NCBI_TaxID=9541;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Kusuda J., Osada N., Hashimoto K.;			
RL	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB083332; BAC20643.1; -.			
FT	NON_TER	1	1	
FT	NON_TER	47	47	
SQ	SEQUENCE	47 AA;	4883 MW;	C5F5537F3A029FFB CRC64;

Query Match 90.4%; Score 85; DB 2; Length 47;
Best Local Similarity 93.8%; Pred. No. 5.8e-07;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1 FSWDNCPEGKDPVIR 16
DB	4 FSWDNCDEGKDPVIR 19

RESULT 2				
Q6LBI5				
ID	Q6LBI5	PRELIMINARY;	PRT;	189 AA.
AC	Q6LBI5;			
DT	05-JUL-2004 (Tremblrel. 27, Created)			
DT	05-JUL-2004 (Tremblrel. 27, Last sequence update)			
DE	05-JUL-2004 (Tremblrel. 27, Last annotation update)			
DE	GM2 activator protein.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92008637; PubMed=1915857;			
RA	Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki			
RA	Sandhoff K.;			
RT	"Characterization of full-length cDNAs and the gene coding for th			
RL	human GM2 activator protein."			
RL	FEBS Lett. 289:260-264(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			

```
RX MEDLINE=93277527; PubMed=8503891;
RA Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA Sandhoff K.;
RT "Over-expression of a functionally active human Gm2-activator protein
RT in escherichia coli.";
RL Biochem. J. 292:571-576(1993).
DR EMBL; X62078; CAA43994.1; -.
DR InterPro; IPR003172; E1_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match          90.4%; Score 85; DB 2; Length 189;
Best Local Similarity 93.8%; Pred. No. 2.7e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVR 16
   ||||| ||||| |||||
Db 30 FSWDNCDEGKDPVAVR 45

RESULT 3
ID CAA43994 PRELIMINARY; PRT; 189 AA.
AC CAA43994;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Gm2 activator protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
RT human Gm2 activator protein.";
RL FEBS Lett. 289:260-264(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277527; PubMed=8503891;
RA Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA Sandhoff K.;
RT "Over-expression of a functionally active human Gm2-activator protein
RT in escherichia coli.";
RL Biochem. J. 292:571-576(1993).
DR EMBL; X62078; CAA43994.1; -.
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match          90.4%; Score 85; DB 2; Length 189;
Best Local Similarity 93.8%; Pred. No. 2.7e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVR 16
   ||||| ||||| |||||
Db 30 FSWDNCDEGKDPVAVR 45

RESULT 4
ID Q8HXX6 PRELIMINARY; PRT; 190 AA.
AC Q8HXX6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Ganglioside Gm2 activator.
GN Name=gma2;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
```

```
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain cerebellum cortex;
RA Kusuda J., Osada N., Hida M., Sugano S., Hashimoto K.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB083313; BAC20592.1; -.
DR HSP; F17900; IG13.
DR InterPro; IPR003172; E1_DerP2_DerF2.
DR SMART; SM00737; ML; 1.
SQ SEQUENCE 190 AA; 20494 MW; 9F9582BEB75715C3 CRC64;

Query Match          90.4%; Score 85; DB 2; Length 190;
Best Local Similarity 93.8%; Pred. No. 2.8e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWDNCFEKGKDPVAVR 16
   ||||| ||||| |||||
Db 31 FSWDNCDEGKDPVAVR 46

RESULT 5
ID SAP3 HUMAN STANDARD; PRT; 193 AA.
AC P17900; O14426; O14428;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Ganglioside Gm2 activator precursor (GM2-AP) (Cerebroside sulfate
DE activator protein) (Shingolipid activator protein 3) (SAP-3).
GN Name=GM2A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RX MEDLINE=91282768; PubMed=2059210;
RA Xie B., McInnes B., Neote K., Lamhonwah A.-M., Mahuran D.;
RT "Isolation and expression of a full-length cDNA encoding the human G-
RT M2 activator protein.";
RL Biochem. Biophys. Res. Commun. 177:1217-1223(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
RT human Gm2 activator protein.";
RL FEBS Lett. 289:260-264(1991).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RC TISSUE=Placenta;
RX MEDLINE=92207171; PubMed=1554364;
RA Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
RT "Evidence for two cDNAs encoding human GM2-activator protein.";
RL Biochem. J. 282:807-813(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93052421; PubMed=1427911;
RA Xie B., Kennedy J.L., McInnes B., Auger D., Mahuran D.J.;
RT "Identification of a processed pseudogene related to the functional
RT gene encoding the GM2 activator protein: localization of the
RT pseudogene to human chromosome 3 and the functional gene to human
RT chromosome 5.";
RL Genomics 14:796-798(1992).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=99294594; PubMed=10364519;
RA Chen B., Rigat B., Curry C., Mahuran D.J.;
RT "Structure of the GM2A gene: identification of an exon 2 nonsense
RT mutation and a naturally occurring transcript with an in-frame
RT deletion of exon 2.";
```

Am. J. Hum. Genet. 65:77-87(1999).

[6]
SEQUENCE FROM N.A.
TISSUE=Uterus;
MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gumaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.R., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez Y., Bouffard G.G.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[7]
SEQUENCE OF 15-193 FROM N.A.
MEDLINE=89325664; PubMed=2753159;
Schroeder M., Klima H., Nakano T., Kwon H., Quintern L.E.,
Gaertner S., Suzuki K., Sandhoff K.;
"Isolation of a cDNA encoding the human GM2 activator protein.";
FEBS Lett. 251:197-200(1989).

[8]
SEQUENCE OF 32-193.
TISSUE=Kidney;
MEDLINE=9106165; PubMed=2209618;
Furst W., Schubert J., Machleidt W., Meyer H.E., Sandhoff K.;
"The complete amino-acid sequences of human ganglioside GM2 activator
protein and cerebroside sulfate activator protein.";
Eur. J. Biochem. 192:709-714(1990).

[9]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
MEDLINE=20545600; PubMed=11090283; DOI=10.1006/jmbi.2000.4225;
Wright C.S., Li S.-C., Rastinejad F.;
"Crystal structure of human GM2-activator protein with a novel beta-
cup topology.";
J. Mol. Biol. 304:411-422(2000).

[10]
VARIANT TSD-AB ARG-138.
MEDLINE=92008638; PubMed=1915858;
Schroeder M., Schnabel D., Suzuki K., Sandhoff K.;
"A mutation in the gene of a glycolipid-binding protein (GM2
activator) that causes GM2-gangliosidosis variant AB.";
FEBS Lett. 290:1-3(1991).

[11]
VARIANT TSD-AB PRO-169.
MEDLINE=94063850; PubMed=8244332;
Schroeder M., Schnabel D., Hurwitz R., Young E., Suzuki K.,
Sandhoff K.;
"Molecular genetics of GM2-gangliosidosis AB variant: a novel mutation
and expression in BHK cells.";
Hum. Genet. 92:437-440(1993).

[12]
VARIANT TSD-AB LYS-88 DEL.
MEDLINE=97055887; PubMed=8900233;
Schepers U., Glombitza G., Lemm T., Hoffmann A., Chabas A., Ozand P.,
Sandhoff K.;
"Molecular analysis of a GM2-activator deficiency in two patients with
GM2-gangliosidosis AB variant.";
Am. J. Hum. Genet. 59:1048-1056(1996).

-I- FUNCTION: Binds gangliosides and stimulates ganglioside GM2
degradation. It stimulates only the breakdown of ganglioside GM2
and glycolipid GA2 by beta-hexosaminidase A. It extracts single

GM2 molecules from membranes and presents them in soluble form to
beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and
conversion to GM3.

-I- SUBCELLULAR LOCATION: Lysosomal.

-I- DISEASE: Defects in GM2A are the cause of Tay-Sachs disease AB
variant (TSD-AB) [MIM:272750]; also known as GM2-gangliosidosis
type AB.

-I- DATABASE: NAME=GM2Adb; NOTE=GM2A mutation database;
WWW="http://www.hexdb.mcgill.ca/?Topic=GM2Adb&Page=MutationSubmission".

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M76477; AAA35907.1; -;
EMBL; X62078; CAA43993.1; -;
EMBL; X61095; CAA43408.1; ALT INIT.
EMBL; L01439; AAA52767.1; -;
EMBL; AF124719; AAD25741.1; -;
EMBL; AF124717; AAD25741.1; JOINED.
EMBL; AF124718; AAD25741.1; JOINED.
EMBL; BC009273; AAH09273.1; -;
EMBL; X16087; CAA34215.1; -;
PIR; I54178; I54178.
PIR; S13195; S13195.
PIR; S22411; S22411.
PDB; 1G13; X-ray; A/B/C=32-193.
Genew; HGNC:4367; GM2A.
MIM; 272750; -;
GO; GO:0005764; C:lysosome; NAS.
GO; GO:0030290; F:sphingolipid activator protein activity; NAS.
GO; GO:0019377; P:glycolipid catabolism; NAS.
GO; GO:0030449; P:sphingolipid catabolism; NAS.
InterPro; IPR003172; E1_DerP2_DerP2.
SMART; SM00737; ML; 1.
3D-structure; Direct protein sequencing; Disease mutation; Signal;
Glycoprotein; GM2-gangliosidosis; Lysosome; Polymorphism; Signal;
Sphingolipid metabolism.
SIGNAL 1 31
FT CHAIN 32 193 Ganglioside GM2 activator.
FT DISULFID 39 183
FT DISULFID 99 106
FT DISULFID 112 138
FT DISULFID 125 136
FT CARBOHYD 63 63
FT VARIANT 19 19 N-linked (GLCNAC...).
FT VARIANT 32 33 /FTid=VAR_013830.
FT VARIANT 88 88 Missing (in 80% of the protein).
FT VARIANT 138 138 /FTid=VAR_006946.
FT VARIANT 169 169 /FTid=VAR_011697.
FT CONFLICT 59 59 C -> R (in TSD-AB).
FT CONFLICT 69 69 R -> P (in TSD-AB).
FT STRAND 35 38 /FTid=VAR_006947.
FT TURN 41 43 R -> I (in TSD-AB).
FT STRAND 46 54 V -> I (in Ref. 3).
FT STRAND 58 59 V -> M (in Ref. 3).
FT STRAND 63 72
FT STRAND 76 76
FT STRAND 81 90
FT STRAND 91 92
FT STRAND 93 96
FT STRAND 100 100
FT STRAND 101 102
FT STRAND 103 103

```

RT STRAND 107 108
FT TURN 109 110
FT HELIX 111 118
FT TURN 121 122

Query Match 90.4%; Score 85; DB 1; Length 193;
Best Local Similarity 93.8%; Pred. No. 2.8e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSWNCFEGKDPVAVR 16
DB 34 FSWNCDCEGKDPVAVR 49

RESULT 6
SAP3 MOUSE
ID SAP3 MOUSE STANDARD; PRT; 193 AA.
AC Q6048; Q61610; Q61819;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate
DE activator protein) (Shingolipid activator protein 3) (SAP-3).
GN Name=GM2a;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=95229165; PubMed=7713516;
RA Yamanaka S., Johnson O.N., Liu M.S., Kozak C.A., Proia R.L.;
RT "The mouse gene encoding the GM2 activator protein (Gm2a): cDNA
RL sequence, expression, and chromosome mapping.";
RL Genomics 24:601-604(1994).
RN [2]_
RP SEQUENCE FROM N.A.
RX MEDLINE=93371367; PubMed=7689829;
RA Bellachioma G., Stirling J.L., Orlicchio A., Beccari T.;
RT "Cloning and sequence analysis of a cDNA clone coding for the mouse
RT GM2 activator protein.";
RL Biochem. J. 294:227-230(1993).
RN [3]_
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA;
RX MEDLINE=97224573; PubMed=9060405;
RA Bertoni C., Appolloni M.G., Stirling J.L., Li S.C., Li Y.T.,
RA Orlicchio A., Beccari T.;
RT "Structural organization and expression of the gene for the mouse GM2
RT activator protein.";
RL Mamm. Genome 8:90-93(1997).
RN [4]_
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

```

"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences".

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

FUNCTION: Binds gangliosides and stimulates ganglioside GM2 degradation. It stimulates only the breakdown of ganglioside GM2 and glycolipid GA2 by beta-hexosaminidase A. It extracts single GM2 molecules from membranes and presents them in soluble form to beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and conversion to GM3.

SUBCELLULAR LOCATION: Lysosomal.

TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney and testis.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U09816; AAA21543.1; -;
EMBL; L19526; AAA61929.1; -;
DR EMBL; U34359; AAB06275.1; ALT SEQ.
DR EMBL; U34356; AAB06275.1; JOINED.
DR EMBL; U34357; AAB06275.1; JOINED.
DR EMBL; U34358; AAB06275.1; JOINED.
DR EMBL; BC004651; AAH04651.1; -;
DR HSSP; P17900; IG13.
DR MGD; MGI:95762; Gm2a.
KW Glycoprotein; Lysosome; Signal; Sphingolipid metabolism.
FT SIGNAL 1 31 By similarity.
FT CHAIN 32 193 Ganglioside GM2 activator.
FT DISULFID 39 183 By similarity.
FT DISULFID 99 106 By similarity.
FT DISULFID 112 138 By similarity.
FT DISULFID 125 136 By similarity.
FT CARBOHYD 151 151 N-linked (GlcNAc...) (Potential).
FT CONFLICT 53 53 I>>T (in Ref.1).
SQ SEQUENCE 193 AA; 20824 MW; 59CC4ABE56FA1FC7 CRC64;
Query Match 87.2%; Score 82; DB 1; Length 193;
Best Local Similarity 87.5%; Pred. No. 9.2e-06;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 FSWNCFEGKDPVAVR 16
DB 34 FSWNCDCEGKDPVAVR 49
RESULT 7
Q6IN37
ID Q6IN37 PRELIMINARY; PRT; 199 AA.
AC Q6IN37;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE GM2 ganglioside activator protein.
GN Name=Gm2a;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gumaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC072474; AAH72474.1; -;
 DR InterPro: IPR003172; E1_DerP2_DerF2.
 DR SMART: SM00737; ML; 1.
 SQ SEQUENCE 199 AA; 21493 MW; C2B5203FDFABF507 CRC64;

Query Match 87.2%; Score 82; DB 2; Length 199;
 Best Local Similarity 87.5%; Pred. No. 9.6e-06;
 Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FSWDNCPEGKDPVAVIR 16
 ||||| |||||
 Db 40 FSWDNCDEGKDPVAVIK 55

RESULT 8
 Q8CJH4

ID Q8CJH4 PRELIMINARY; PRT; 199 AA.
 AC Q8CJH4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE GM2 activator protein.
 GN Name=GM2AP;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Miwa N., Okada T., Nakamura S.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB051391; BAC24018.1; -;
 DR HSSP; P17900; IGL13.
 DR InterPro: IPR003172; E1_DerP2_DerF2.
 DR SMART; SM00737; ML; 1.
 SQ SEQUENCE 199 AA; 21521 MW; D585203FDFABF507 CRC64;

Query Match 87.2%; Score 82; DB 2; Length 199;
 Best Local Similarity 87.5%; Pred. No. 9.6e-06;
 Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FSWDNCPEGKDPVAVIR 16
 ||||| |||||
 Db 40 FSWDNCDEGKDPVAVIK 55

RESULT 9
 Q9QJW2

ID Q9QJW2 PRELIMINARY; PRT; 20 AA.
 AC Q9QJW2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 22 kDa INTRALYSOSOMAL lipid-transfer protein (Fragment).

OS Rattus sp.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10118;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=97104296; PubMed=8948454;
 RA Kuwana T., Mullock B.M., Luzzio J.P.;
 RT "Identification of a lysosomal protein causing lipid transfer, using a
 RT fluorescence assay designed to monitor membrane fusion between rat
 RT liver endosomes and lysosomes.";
 RL Biochem. J. 308:937-946(1995).
 DR HSSP; P17900; IGL13.
 SQ SEQUENCE 20 AA; 2211 MW; 3A160591EA5542C5 CRC64;

Query Match 63.8%; Score 60; DB 2; Length 20;
 Best Local Similarity 75.0%; Pred. No. 0.0045;
 Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FSWDNCPEGKDPVAVIR 16
 ||||| |||||
 Db 1 FSWDNCDEGKDPVAVIK 16

RESULT 10
 Q8CDW5

ID Q8CDW5 PRELIMINARY; PRT; 338 AA.
 AC Q8CDW5;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched
 DE library, clone:4833441N19 product:hypothetical Prokaryotic membrane
 DE lipoprotein lipid attachment site/NHL repeat containing protein, full
 DE insert sequence. (Fragment).
 GN Name=8030451K01Rik;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.

OrderedLocusNames=bl17607;
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Bradyrhizobiaceae; Bradyrhizobium.
 RN NCBI_TaxID=375;
 [1]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN=USDA110;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
 RA Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 Bradyrhizobium japonicum USDA110.";
 RL DNA Res. 9:189-197(2002).
 CC - SIMILARITY: Belongs to the aldehyde dehydrogenase family.
 DR EMBL: AP005962; BAC52872.1; -.
 DR HSSP: P05091; 1002.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.
 DR InterPro: IPR002086; Aldehyde dehydr.
 DR Pfam: PF00171; Aldehyde; 1.
 DR PROSITE: PS00070; ALDEHYDE DEHYDR. CYS; UNKNOWN 1.
 DR PROSITE: PS00687; ALDEHYDE DEHYDR. GLU; 1.
 KW Complete proteome; Oxidoreductase.
 SQ SEQUENCE 565 AA; 60599 MW; 784E16D8AE7622B2 CRC64;
 Query Match 51.1%; Score 48; DB 2; Length 565;
 Best Local Similarity 58.3%; Pred. No. 22;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 3 WDNCFCGKDPVAV 14
 Db 521 WVNQYQMDPAV 532
 ID Q6BY19 PRELIMINARY; PRT; 2802 AA.
 AC Q6BY19; 2802 AA.
 DT 01-OCT-2004 (TrEMBLrel. 28, Created)
 DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Chromosome A of strain CBS767 of Debaryomyces hansenii.
 GN ORFNames=DEHA0409427g;
 OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
 RN NCBI_TaxID=4959;
 [1]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN=CBS767;
 RG GENOLEVURES;
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
 RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
 RA Boissrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaud J.M., Nikolski M., Ostas S., Olier-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
 RA Swennene D., Tekata F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
 RA Wincker P., Souciet J.L.;
 RT "Genome evolution in yeasts.";
 RL Nature 430:35-44(2004).
 [2]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN=CBS767;

RA Genoscope;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: CR382133; CAG84689.1; -.
 SQ SEQUENCE 2802 AA; 317793 MW; C8447351E3CB91BA CRC64;
 Query Match 51.1%; Score 48; DB 2; Length 2802;
 Best Local Similarity 43.8%; Pred. No. 1.3e+02;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
 QY 1 FSWDNCFCGKDPVAVR 16
 Db 1895 FIMNCLGKPLMKV 1910
 ID Q9YB02 PRELIMINARY; PRT; 171 AA.
 AC Q9YB02;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein APE1793.
 GN OrderedLocusNames=APE1793;
 OS Aeropyrum pernix.
 CC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 CC Desulfurococaceae; Aeropyrum.
 RN NCBI_TaxID=56636;
 [1]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382986;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL: AP000062; BAA80796.1; -.
 DR PIR: G72563; G72563.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 171 AA; 17891 MW; 3A981F6212B12D63 CRC64;
 Query Match 50.0%; Score 47; DB 2; Length 171;
 Best Local Similarity 37.5%; Pred. No. 8.5;
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 QY 1 FSWDNCFCGKDPVAVR 16
 Db 44 WGNWSCFSGLKPLMR 59
 ID Q7T362 PRELIMINARY; PRT; 282 AA.
 AC Q7T362;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Nucleophosmin 1.
 GN Names=nmpl;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 CC Cyprinidae; Danio.
 RN NCBI_TaxID=7955;
 [1]
 RN RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Kidney;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053240; AAH53240.1; -.
DR ZFIN; ZDB-GENE-021028-1; npml.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004301; Nucleoplasm.
DR Pfam; PF03066; Nucleoplasm; 1.
SQ SEQUENCE 282 AA; 31490 MW; AD11D8E097F9F1E9 CRC64;

Query Match 50.0%; Score 47; DB 2; Length 282;
Best Local Similarity 58.3%; Pred.No.15;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 DMCFGKDPFVI 15
Db :|||:|:|:
260 NNCFGTDPKV 271

Search completed: November 17, 2004, 07:22:35
Job time : 6.63957 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 09:12:54 ; Search time 4.68912 Seconds
(without alignments)
1224.039 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 16

Sequence: 1 YSLPKSEFAVDLELP 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq_23Sep04.*

- 1: geneseqp1980s.*
- 2: geneseqp1990s.*
- 3: geneseqp2000s.*
- 4: geneseqp2001s.*
- 5: geneseqp2002s.*
- 6: geneseqp2003as.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	16	4	AAB31927
2	16	100.0	193	4	AAB31897
3	8	50.0	178	4	AAB31898
4	8	50.0	178	5	ABG31346
5	8	50.0	189	4	AAB31900
6	8	50.0	193	4	AAB31901
7	8	50.0	193	4	AAB31902
8	8	50.0	193	4	AAB31904
9	8	50.0	193	4	AAB31896
10	8	50.0	193	4	AAB31928
11	8	50.0	193	4	AAB31903
12	8	50.0	193	4	ABG00720
13	8	50.0	193	5	ABG31345
14	8	50.0	193	5	ABG31345
15	8	50.0	193	7	ADN95858
16	8	50.0	193	8	ADN95858
17	8	50.0	193	8	ADN03620
18	8	50.0	200	4	ADQ17712
19	7	43.8	199	2	AAW10656
20	7	43.8	534	6	ABM69301
21	6	37.5	39	4	AAE03304
22	6	37.5	39	4	AAE03334
23	6	37.5	39	5	ABG64473
24	6	37.5	39	5	ABG64475
25	6	37.5	39	8	ADL77740

RESULT 1
AAB31927
ID AAB31927 standard; peptide; 16 AA.
XX AC AAB31927;
XX DT 15-MAY-2001 (first entry)
XX DE Amino acid sequence of a peptide fragment of a human protein.
XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
XX KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
XX KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
XX KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
XX KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX OS Homo sapiens.
XX EN WO200105422-A2.
XX PD 25-JAN-2001.
XX PF 17-JUL-2000; 2000WO-FR002057.
XX PR 15-JUL-1999; 99FR-00009372.
XX PA (INMR) BIOMERIEUX STELHYS.
XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX DR WPI; 2001-159475/16.
XX DT Detecting, preventing and treating degenerative, neurological and
XX PT autoimmune diseases, particularly multiple sclerosis, using specified
XX PT polypeptides or related nucleic acid or ligand.
XX PS Claim 24; Page 108; 209pp; French.
XX CC The present sequence represents a peptide fragment of a human protein,
XX CC which is used in the method of the invention. The specification describes
XX CC a method which uses at least one polypeptide or polynucleotide sequence
XX CC belonging to the perlecan, precursor of the retinol-binding plasma
XX CC protein, precursor of the ganglioside GM2 activator, calgranulin B or
XX CC saposin B protein families. The method is used for detecting, preventing
XX CC or treating a degenerative, neurological and/or auto-immune disease. The
XX CC polynucleotides and polypeptides are used for diagnosis, prognosis, and
XX CC prevention and treatment of multiple sclerosis (in its various forms and
XX CC phases). They may also be useful in cases of e.g. Alzheimer's and

ALIGNMENTS

26	6	37.5	39	8	ADL77742	AdL77742 Albumin f
27	6	37.5	40	4	AAU01938	Aau01938 Human sec
28	6	37.5	54	5	ABP05760	Abp05760 Human ORF
29	6	37.5	57	5	ABP00255	Abp00255 Human ORF
30	6	37.5	95	3	AA53954	Aa53954 Human col
31	6	37.5	101	6	ABU18128	Abu18128 Protein e
32	6	37.5	112	7	ABO67767	AbO67767 Klebsiell
33	6	37.5	122	7	ADD12565	Add12565 Human ENZ
34	6	37.5	138	3	AA50654	Aa50654 Arabidops
35	6	37.5	151	4	AA91093	Aa91093 C Glutami
36	6	37.5	160	3	AA50653	Aa50653 Arabidops
37	6	37.5	176	4	AB84331	Ab84331 Amino aci
38	6	37.5	176	4	AAW78987	AaW78987 Human pro
39	6	37.5	176	7	ADJ70046	AdJ70046 Human hea
40	6	37.5	183	4	AAW79971	AaW79971 Human pro
41	6	37.5	201	3	AA41811	Aa41811 Human ORF
42	6	37.5	214	2	AAV31625	AaV31625 Human cal
43	6	37.5	214	6	ABU41132	Abu41132 Protein e
44	6	37.5	214	7	ADE59915	Ade59915 Rat Prote
45	6	37.5	214	7	ADE59919	Ade59919 Rat Prote

CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 16 AA;

Query Match 100.0%; Score 16; DB 4; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
 |||||
 DB 1 YSLPKSEFAVPDLELP 16

RESULT 2
 AAB31897
 ID AAB31897 standard; protein; 193 AA.

AC AAB31897;

DT 15-MAY-2001 (first entry)

XX Amino acid sequence of a mutant ganglioside GM2 activator protein.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX Homo sapiens.

XX WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INMR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX Claim 25; Page 159-160; 209pp; French.

XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 193 AA;

Query Match 100.0%; Score 16; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
 |||||
 DB 145 YSLPKSEFAVPDLELP 160

RESULT 3
 AAB31898
 ID AAB31898 standard; protein; 178 AA.

XX AAB31898;

XX 15-MAY-2001 (first entry)

XX Amino acid sequence of a human protein.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX Homo sapiens.

XX WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INMR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX Claim 1; Page 160; 209pp; French.

XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 178 AA;

Query Match 50.0%; Score 8; DB 4; Length 178;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEF 8
 |||||
 DB 130 YSLPKSEF 137

RESULT 4
ABG31346
ID ABG31346 standard; protein; 178 AA.
XX AC ABG31346;
XX DT 15-NOV-2002 (first entry)
XX DE Non-glycosylated human GM2 activator protein containing His6-tag.
XX KW Human; GM2 activator protein; ganglioside; platelet activating factor;
KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
KW antiinflammatory; antiasthmatic; antiarthritic; antirheumatic;
KW antiallergic; hepatotropic; nephrotropic; immunosuppressive;
KW tranquilizer.
XX OS Homo sapiens.
OS Synthetic.
XX FH Key Location/Qualifiers
FT Binding-site 5..10
FT /label= His6 tag
FT /note= "Nickel chelating region used for purification of
FT the protein"
FT Cleavage-site 11..16
FT /label= Factor_X_cleavage_site
XX US6423680-B1.
XX 23-JUL-2002.
XX 30-OCT-1998; 98US-00183841.
XX 30-OCT-1998; 98US-00183841.
XX (HSCR-) HSC RES & DEV LP.
XX Rigat B, Reynaud D, Mahuran D;
XX WPI; 2002-664636/71.
XX Composition useful for treating inflammatory conditions e.g. asthma
XX comprises GM2 activator protein or GM2 activator peptide in combination
XX with a carrier.
XX Example 1; Fig 2; 11pp; English.
XX The present invention relates to a composition comprising GM2
XX (ganglioside) activator protein or a GM2 activator peptide derived from
XX the GM2 activator protein in combination with a carrier. The composition
XX comprises the protein or peptide in an amount of 1-100 mg. The GM2
XX activator protein is capable of inhibiting platelet activating factor
XX (PAF). The composition of the invention is useful for treating
XX inflammatory disorders e.g. inflammatory bowel disease, asthma,
XX autoimmune disease (such as lupus), hypersensitivity infection,
XX rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
XX gout and tissue-specific conditions (e.g. glomerulonephritis and
XX hepatitis). The composition is capable of inhibiting platelet activating
XX factor, is non-toxic, is efficacious and presents less severe side
XX effects, including redness, swelling, pain and polymorphonuclear
XX leukocyte accumulation at the inflammatory site and other associated
XX cellular responses. The present sequence represents a non-glycosylated
XX human GM2 activator protein prepared using a His6-tag bacterial
XX expression system
XX SQ Sequence 178 AA;
Query Match 50.0%; Score 8; DB 5; Length 178;

Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YSLPKSEF 8
DB 130 YSLPKSEF 137
|||||
RESULT 5
AAB31900
ID AAB31900 standard; protein; 189 AA.
XX AC AAB31900;
XX DT 15-MAY-2001 (first entry)
XX DE Amino acid sequence of a human protein.
XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX OS Homo sapiens.
XX WO200105422-A2.
XX 25-JAN-2001.
XX 17-JUL-2000; 2000WO-FR002057.
XX 15-JUL-1999; 99FR-00009372.
XX (INMR) BIOMERIEUX STELHYS.
XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand.
XX Claim 1; Page 161-162; 209pp; French.
XX The present sequence represents a human protein, which is used in the
XX method of the invention. The specification describes a method which uses
XX at least one polypeptide or polynucleotide sequence belonging to the
XX perlecan, precursor of the retinol-binding plasma protein, precursor of
XX the ganglioside GM2 activator, calgranulin B or saposin B protein
XX families. The method is used for detecting, preventing or treating a
XX degenerative, neurological and/or auto-immune disease. The
XX polynucleotides and polypeptides are used for diagnosis, prognosis,
XX prevention and treatment of multiple sclerosis (in its various forms and
XX phases). They may also be useful in cases of e.g. Alzheimer's and
XX Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
XX polyarthritis and lupus erythematosus, including use as vaccines and in
XX gene therapy (expression of sense or antisense sequences). They can also
XX be used to assess efficacy of potential therapeutic agents, particularly
XX compounds that reduce or inhibit toxicity towards glial cells
XX SQ Sequence 189 AA;
Query Match 50.0%; Score 8; DB 4; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YSLPKSEF 8
DB 141 YSLPKSEF 148
|||||

```

RESULT 6
ID AAB31901 standard; protein; 193 AA.
XX
AC AAB31901;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX
DR WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 162-163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 50.0%; Score 8; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YSLPKSEF 8
Db 145 YSLPKSEF 152
RESULT 7
ID AAB31902 standard; protein; 193 AA.
XX
AC AAB31902;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX
DR WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 162-163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 50.0%; Score 8; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YSLPKSEF 8
Db 145 YSLPKSEF 152

```


XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX Disclosure; Fig 1; 209pp; French.
 XX The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX SQ Sequence 193 AA;

Query Match 50.0%; Score 8; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEF 8
 |||||
 DB 145 YSLPKSEF 152

RESULT 11
 AAB31903
 ID AAB31903 standard; protein; 193 AA.
 XX AAB31903;
 XX 15-MAY-2001 (first entry)
 XX Amino acid sequence of a human protein.
 XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 XX ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 XX neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 XX Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 XX rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX Homo sapiens.
 XX WO200105422-A2.
 XX 25-JAN-2001.
 XX 17-JUL-2000; 2000WO-FR002057.
 XX 15-JUL-1999; 99FR-00009372.
 XX (INMR) BIOMERIEUX STELHYS.
 XX Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX Claim 1; Page 164; 209pp; French.
 XX The present sequence represents a human protein, which is used in the

CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX SQ Sequence 193 AA;

Query Match 50.0%; Score 8; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEF 8
 |||||
 DB 145 YSLPKSEF 152

RESULT 12
 ABG00720
 ID ABG00720 standard; protein; 193 AA.
 XX ABG00720;
 XX 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #711.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 XX WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX N-PSDB; AAS64907.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID NO 31079; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: the sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 193 AA;

Query Match 50.0%; Score 8; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEF 8
 |||||
 Db 145 YSLPKSEF 152

RESULT 13

ABG31345
 ID ABG31345 standard; protein; 193 AA.

XX AC ABG31345;

XX DT 15-NOV-2002 (first entry)

XX DE Human GM2 activator protein.

XX KW Human; GM2 activator protein; ganglioside; platelet activating factor;
 KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
 KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
 KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
 KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
 KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
 KW antiinflammatory; antiasthmatic; antiarthritic; antirheumatic;
 KW antiallergic; hepatotropic; nephrotropic; immunosuppressive;
 KW tranquilliser.

XX OS Homo sapiens.

XX PN US6423680-B1.

XX PD 23-JUL-2002.

XX PF 30-OCT-1998; 98US-00183841.

XX PR 30-OCT-1998; 98US-00183841.

XX PA (HSCR-) HSC RES & DEV LP.

XX PI Rigat B, Reynaud D, Mahuran D;

XX DR WPI; 2002-664636/71.

XX PT Composition useful for treating inflammatory conditions e.g. asthma
 PT comprises GM2 activator protein or GM2 activator peptide in combination
 PT with a carrier.

XX PS Claim 3; Fig 1; 1lpp; English.

XX CC The present invention relates to a composition comprising GM2
 CC (ganglioside) activator protein or a GM2 activator peptide derived from
 CC the GM2 activator protein in combination with a carrier. The composition
 CC comprises the protein or peptide in an amount of 1-100 mg. The GM2
 CC activator protein is capable of inhibiting platelet activating factor
 CC (PAF). The composition of the invention is useful for treating
 CC inflammatory disorders e.g. inflammatory bowel disease, asthma,

CC autoimmune disease (such as lupus), hypersensitivity infection,
 CC rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
 CC gout and tissue-specific conditions (e.g. glomerulonephritis and
 CC hepatitis). The composition is capable of inhibiting platelet activating
 CC factor, is non-toxic, is efficacious and presents less severe side
 CC effects, including redness, swelling, pain and polymorphonuclear
 CC leukocyte accumulation at the inflammatory site and other associated
 CC cellular responses. The present sequence represents human GM2 activator
 CC protein

XX Sequence 193 AA;

Query Match 50.0%; Score 8; DB 5; Length 193;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEF 8

|||||
 Db 145 YSLPKSEF 152

RESULT 14

ABP65212
 ID ABP65212 standard; protein; 193 AA.

XX AC ABP65212;

XX DT 12-NOV-2002 (first entry)

XX DE Hypoxia-regulated protein #86.

XX KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;
 KW antiinflammatory; vulnery; gynecological; ophthalmological; vaccine;
 KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
 KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;
 KW preclapsmia; atherosclerosis; inflammatory condition; wound healing;
 KW inflammation; erythropoiesis; hair loss; human.

XX OS Homo sapiens.

XX PN WO200246465-A2.

XX PD 13-JUN-2002.

XX PF 10-DEC-2001; 2001WO-GB005458.

XX PR 08-DEC-2000; 2000GB-00030076.

XX PR 08-FEB-2001; 2001GB-00003156.

XX PR 25-OCT-2001; 2001GB-00025666.

XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX PI White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;

XX PI Rayner WN;

XX DR WPI; 2002-627238/67.

XX PT Identifying a gene involved in disease for treating hypoxia-regulated
 PT conditions, comprises comparing the transcriptome/proteome of two cell
 PT types under different conditions and identifying a differentially
 PT regulated gene.

XX PS Claim 35; Page 397; 538pp; English.

XX CC The present invention relates to methods for identifying genes and
 CC proteins that are implicated in a specific disease or physiological
 CC condition. The method comprises comparing the transcriptome/proteome of a
 CC second specialised cell type implicated in a disease or condition with that of a
 CC second specialised cell type, under two experimental conditions, and
 CC identifying a gene that is differentially regulated in the two
 CC specialised cell types under experimental conditions. ABV7873-ABV78116
 CC and ABP65061-ABP65257 were identified using the methods of the invention.
 CC The coding sequences and proteins are useful for treating a disease in a

CC patient, for manufacture of a medicament for treating hypoxia-regulated
 CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
 CC biological response to hypoxic conditions, or hypoxic-associated
 CC pathology in a patient. The coding sequences and proteins are also useful
 CC for monitoring the therapeutic treatment of a disease or physiological
 CC condition, such as cancer, ischaemic conditions, reperfusion injury,
 CC retinopathy, neonatal stress, pre-eclampsia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss
 XX Sequence 193 AA;

Query Match 50.0%; Score 8; DB 5; Length 193;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEF 8
 |||||
 Db 145 YSLPKSEF 152

RESULT 15

ADN95858
 ID ADN95858 standard; protein; 193 AA.

XX AC
 XX ADN95858;

XX DT 01-JUL-2004 (first entry)

XX DE Human BEC/LEC-related protein sequence SeqID782.

XX KW growth; differentiation; blood endothelial cell; BEC;
 KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
 KW lymphatic growth agent; VEGF-C; VEGF-D; angiogenic; cytostatic;
 KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
 KW inflammatory disease; cancer metastasis; lymphatic system; human.

XX OS Homo sapiens.

XX PN WO2003080640-A1.

XX PD 02-OCT-2003.

XX PF 07-MAR-2003; 2003WO-US006900.

XX PR 07-MAR-2002; 2002US-0363019P.

XX PA (LUDW-) LUDWIG INST CANCER RES.
 XX PA (LICN) LICENTIA LTD.

XX PI Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;

XX DR WPI; 2003-876899/81.

XX DR N-PSDB; ADN95859.

XX PS Example 1; SEQ ID NO 782; 176pp; English.

CC This invention relates to a method of differentially modulating the
 CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
 CC endothelial cells (LEC) comprising contacting endothelial cells with a
 CC composition comprising an agent that differentially modulates blood or
 CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
 CC identifying a human subject with lymphoedema and with a mutation in at
 CC least one allele of a gene encoding a LEC protein, where the mutation
 CC correlates with lymphoedema in human subjects, and with the proviso that
 CC the LEC protein is not VEGFR-3; and administering to the subject a
 CC composition comprising a lymphatic growth agent selected from VEGF-C or
 CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
 CC the development of compounds with an antiangiogenic, cytostatic,
 CC vasotropic or antiinflammatory activity or for gene therapy. The method
 CC is useful in modulating the growth or differentiation of blood
 CC endothelial cells or lymphatic endothelial cells, in treating hereditary
 CC lymphoedema, in screening for an endothelial cell disorder or
 CC predisposition to the disorder or in monitoring the efficacy or toxicity

CC of a drug on endothelial cells. The agent is useful in manufacturing a
 CC medicament for the differential modulation of blood vessel endothelial
 CC cell or lymphatic vessel endothelial cell growth or differentiation. The
 CC lymphatic growth agent may also be used in manufacturing a medicament for
 CC the treatment of hereditary lymphoedema resulting from a mutation in a
 CC LEC gene or of other diseases involving the lymphatic vessels, such as
 CC various inflammatory diseases and cancer metastasis via the lymphatic
 CC system. The present sequence is that of a human LEC/BEC differentially
 CC expressed protein which is related to the method of the invention. Note:
 CC This sequence does not appear in the specification but was obtained by
 CC the indexer using the source data given in table 14 of the specification.

XX SQ Sequence 193 AA;

Query Match 50.0%; Score 8; DB 7; Length 193;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEF 8

|||||

Db 145 YSLPKSEF 152

Search completed: November 17, 2004, 10:33:01

Job time : 6.78003 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 10:53:27 : Search time 1.14595 Seconds
(without alignments)
925.943 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 16

Sequence: 1 YSLPKSEFAVFDLELP 16

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pbp:*
- 2: /cgn2_6/ptodata/1/iaa/5B COMB.pbp:*
- 3: /cgn2_6/ptodata/1/iaa/6A COMB.pbp:*
- 4: /cgn2_6/ptodata/1/iaa/6B COMB.pbp:*
- 5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pbp:*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pbp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	50.0	178	4	US-09-183-841-2
2	8	50.0	193	4	US-09-183-841-1
3	6	37.5	55	4	US-09-270-767-62426
4	6	37.5	112	4	US-09-489-039A-14284
5	6	37.5	219	4	US-09-248-796A-16515
6	6	37.5	223	4	US-09-543-681A-7057
7	6	37.5	253	4	US-09-583-110-3681
8	6	37.5	255	4	US-08-778-717-11
9	6	37.5	257	4	US-08-778-717-13
10	6	37.5	258	4	US-09-543-681A-8135
11	6	37.5	329	4	US-09-719-108-6
12	6	37.5	349	4	US-09-270-767-42023
13	6	37.5	356	4	US-09-270-767-46804
14	6	37.5	473	4	US-09-252-991A-23441
15	6	37.5	534	4	US-09-134-000C-5087
16	6	37.5	537	3	US-08-886-886-17
17	6	37.5	547	4	US-09-187-532A-5905
18	6	37.5	580	4	US-09-489-039A-14205
19	6	37.5	834	4	US-09-252-991A-17616
20	6	37.5	2329	3	US-08-755-587-16
21	6	37.5	2618	3	US-09-413-814-28
22	6	37.5	3418	2	US-08-639-501-2
23	6	37.5	3418	2	US-08-603-753D-4
24	6	37.5	3418	3	US-09-044-946-2
25	6	37.5	3418	3	US-08-755-587-44
26	6	37.5	3418	3	US-09-044-908-2
27	6	37.5	3418	3	US-09-099-753-4

```

28 6 37.5 3418 3 US-08-986-106-4 Sequence 4, Appli
29 5 31.2 6 1 US-08-467-607-12 Sequence 12, Appl
30 5 31.2 6 2 US-08-469-362-12 Sequence 12, Appl
31 5 31.2 6 2 US-08-850-392-12 Sequence 12, Appl
32 5 31.2 11 3 US-09-476-482-11 Sequence 11, Appl
33 5 31.2 15 3 US-08-946-026-57 Sequence 57, Appl
34 5 31.2 19 2 US-08-793-490-5 Sequence 5, Appli
35 5 31.2 20 1 US-08-467-607-4 Sequence 4, Appli
36 5 31.2 20 2 US-08-469-362-4 Sequence 4, Appli
37 5 31.2 20 2 US-08-850-392-4 Sequence 4, Appli
38 5 31.2 21 3 US-09-406-781-40 Sequence 40, Appli
39 5 31.2 21 4 US-09-880-132-40 Sequence 40, Appli
40 5 31.2 35 3 US-09-079-372-11 Sequence 11, Appli
41 5 31.2 48 4 US-09-205-258-574 Sequence 574, App
42 5 31.2 56 4 US-09-270-767-61615 Sequence 61615, A
43 5 31.2 59 4 US-09-010-147B-8 Sequence 8, Appli
44 5 31.2 61 4 US-09-621-976-5270 Sequence 5270, Ap
45 5 31.2 61 4 US-09-248-796A-24265 Sequence 24265, A

```

ALIGNMENTS

```

RESULT 1
US-09-183-841-2
; Sequence 2, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: His tag at residues 1 to 17
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: amino acid
; OTHER INFORMATION: sequence of GM2 protein using His6 tag
US-09-183-841-2

```

```

Query Match 50.0%; Score 8; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 YSLPKSEF 8
    |||||
Db 130 YSLPKSEF 137

```

```

RESULT 2
US-09-183-841-1
; Sequence 1, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL

```

```

; LOCATION: (33)...(55)
; FEATURE:
; OTHER INFORMATION: residues 56-63 are included in a further precursor
; OTHER INFORMATION: form of the protein
US-09-183-841-1

Query Match          50.0%; Score 8; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEF 8
Db 145 YSLPKSEF 152

RESULT 3
US-09-270-767-62426
; Sequence 62426, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62426
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-62426

Query Match          37.5%; Score 6; DB 4; Length 55;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLPKSE 7
Db 32 SLPKSE 37

RESULT 4
US-09-489-039A-14284
; Sequence 14284, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 14284
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14284

Query Match          37.5%; Score 6; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FAVPDL 13
Db 107 FAVPDL 112

```

```

RESULT 5
US-09-248-796A-16515
; Sequence 16515, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16515
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16515

Query Match          37.5%; Score 6; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KSEFAV 10
Db 207 KSEFAV 212

RESULT 6
US-09-543-681A-7057
; Sequence 7057, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7057
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7057

Query Match          37.5%; Score 6; DB 4; Length 223;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 EFAVPD 12
Db 117 EFAVPD 122

RESULT 7
US-09-583-110-3681
; Sequence 3681, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30

```

```

; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3681
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-3681

Query Match      37.5%; Score 6; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 KSEFAV 10
Db      43 KSEFAV 48

RESULT 8
US-08-778-717-11
; Sequence 11, Application US/08778717
; Patent No. 6602689
; GENERAL INFORMATION:
; APPLICANT: UENO, EIICHI
; APPLICANT: NOBUYUKI, FUJII
; APPLICANT: OKADA, MASAHISA
; TITLE OF INVENTION: FUSED DNA SEQUENCE, FUSED PROTEIN
; TITLE OF INVENTION: EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR
; TITLE OF INVENTION: EXPRESSING SAID DNA SEQUENCE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/778,717
; FILING DATE: 12-DEC-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION NUMBER: JP 352225/1995
; FILING DATE: 28-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2084-031-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 255 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: RECOMBINANT
; PUBLICATION INFORMATION:
; AUTHORS: NOBUYUKI FUJII ET AL,
; TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED
; TITLE: FROM SAID FUSED DNA SEQUENCE AND METHOD OF
; TITLE: EXPRESSING SAID FUSED PROTEIN

```

```

; RELEVANT RESIDUES IN SEQ ID NO: 11: FROM 1 TO 255
US-08-778-717-11

Query Match      37.5%; Score 6; DB 4; Length 255;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 SEFAVP 11
Db      77 SEFAVP 82

RESULT 9
US-08-778-717-13
; Sequence 13, Application US/08778717
; Patent No. 6602689
; GENERAL INFORMATION:
; APPLICANT: UENO, EIICHI
; APPLICANT: NOBUYUKI, FUJII
; APPLICANT: OKADA, MASAHISA
; TITLE OF INVENTION: FUSED DNA SEQUENCE, FUSED PROTEIN
; TITLE OF INVENTION: EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR
; TITLE OF INVENTION: EXPRESSING SAID DNA SEQUENCE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/778,717
; FILING DATE: 12-DEC-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION NUMBER: JP 352225/1995
; FILING DATE: 28-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2084-031-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: RECOMBINANT
; PUBLICATION INFORMATION:
; AUTHORS: NOBUYUKI FUJII ET AL,
; TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED
; TITLE: FROM SAID FUSED DNA SEQUENCE AND METHOD OF
; TITLE: EXPRESSING SAID FUSED PROTEIN
; RELEVANT RESIDUES IN SEQ ID NO: 13: FROM 1 TO 257
US-08-778-717-13

Query Match      37.5%; Score 6; DB 4; Length 257;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 SEFAVP 11

```

```
Db      77 SEFAVP 82
|||||
RESULT 10
US-09-543-681A-8135
; Sequence 8135, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8135
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8135

Query Match      37.5%; Score 6; DB 4; Length 258;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 SLPKSE 7
|||||
Db      165 SLPKSE 170

RESULT 11
US-09-719-108-6
; Sequence 6, Application US/09719108
; Patent No. 6670527
; GENERAL INFORMATION:
; APPLICANT: Thomas, Stephen G
; APPLICANT: Hedden, Peter
; APPLICANT: Phillips, Andrew L
; TITLE OF INVENTION: Gibberellin 2-Oxidase
; FILE REFERENCE: 0623.0970000
; CURRENT APPLICATION NUMBER: US/09/719,108
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: PCT/GB99/01857
; PRIOR FILING DATE: 1999-06-11
; PRIOR APPLICATION NUMBER: GB 9812821.8
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: GB 9815404.0
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-719-108-6

Query Match      37.5%; Score 6; DB 4; Length 329;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 SLPKSE 7
|||||
Db      66 SLPKSE 71

RESULT 12
US-09-270-767-42023
; Sequence 42023, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
```

```
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42023
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-42023

Query Match      37.5%; Score 6; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11 PDLELP 16
|||||
Db      200 PDLELP 205

RESULT 13
US-09-270-767-46804
; Sequence 46804, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46804
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-46804

Query Match      37.5%; Score 6; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 SLPKSE 7
|||||
Db      333 SLPKSE 338

RESULT 14
US-09-252-991A-23441
; Sequence 23441, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23441
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23441
```

Query Match 37.5%; Score 6; DB 4; Length 473;
 Best Local Similarity 100.0%; Pred.No.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 AVPDLE 14
 Db 256 AVPDLE 261

RESULT 15
 US-09-134-000C-5087
 ; Sequence 5087; Application US/09134000C
 ; Patent No. 6617156
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; FILE REFERENCE: 032796-032
 ; CURRENT APPLICATION NUMBER: US/09/134.000C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/055,778
 ; PRIOR FILING DATE: 1997-08-15
 ; NUMBER OF SEQ ID NOS: 6812
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 5087
 ; LENGTH: 534
 ; TYPE: PRT
 ; ORGANISM: Enterococcus faecalis
 US-09-134-000C-5087

Query Match 37.5%; Score 6; DB 4; Length 534;
 Best Local Similarity 100.0%; Pred.No.1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 SLPKSE 7
 Db 1 SLPKSE 6

Search completed: November 17, 2004, 12:29:33
 Job time : 2.23686 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: November 17, 2004, 12:24:02 ; Search time 3.8405 Seconds
(without alignments)
1475.341 Million cell updates/sec

Title: US-10-030-937-72
Perfect score: 16
Sequence: 1 YSLPKSEFAVPDLELP 16
Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 1570615 seqs, 354127592 residues
Word size : 0
Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	8	50.0	193	US-10-170-385-389
2	7	43.8	54	Sequence 389, App
3	6	37.5	17	Sequence 220040,
4	6	37.5	39	Sequence 1222, Ap
5	6	37.5	11	Sequence 1224, Ap
6	6	37.5	51	Sequence 105909,
7	6	37.5	71	Sequence 252984,
8	6	37.5	86	Sequence 188997,
9	6	37.5	92	Sequence 247004,
10	6	37.5	95	US-10-424-599-247004
11	6	37.5	95	US-09-925-299-1494
12	6	37.5	101	Sequence 1494, Ap
13	6	37.5	116	Sequence 46052, A
	6	37.5	117	Sequence 188964,
	6	37.5	117	Sequence 358214,

14	6	37.5	120	15	US-10-424-599-284187	Sequence 284187,
15	6	37.5	127	16	US-10-437-963-190106	Sequence 190106,
16	6	37.5	129	16	US-10-437-963-133889	Sequence 133889,
17	6	37.5	131	15	US-10-425-114-57104	Sequence 57104, A
18	6	37.5	151	9	US-09-738-626-4847	Sequence 4847, Ap
19	6	37.5	157	17	US-10-425-115-269204	Sequence 269204, A
20	6	37.5	165	15	US-10-425-114-64412	Sequence 64412, A
21	6	37.5	176	14	US-10-168-066-6	Sequence 6, Appli
22	6	37.5	176	16	US-10-408-765A-1852	Sequence 1852, Ap
23	6	37.5	202	16	US-10-767-701-54502	Sequence 54502, A
24	6	37.5	214	9	US-09-999-602-1	Sequence 1, Appli
25	6	37.5	214	15	US-10-282-122A-69056	Sequence 69056, A
26	6	37.5	216	14	US-10-301-822-217	Sequence 217, App
27	6	37.5	216	15	US-10-425-114-50218	Sequence 50218, A
28	6	37.5	216	16	US-10-734-564-111	Sequence 111, App
29	6	37.5	227	17	US-10-425-115-255964	Sequence 255964,
30	6	37.5	249	15	US-10-425-114-69119	Sequence 69119, A
31	6	37.5	251	14	US-10-301-822-220	Sequence 220, App
32	6	37.5	255	15	US-10-457-372-11	GENERAL INFORMA
33	6	37.5	257	14	US-10-106-698-6128	Sequence 6128, Ap
34	6	37.5	257	15	US-10-457-372-13	GENERAL INFORMA
35	6	37.5	267	14	US-10-301-822-215	Sequence 215, App
36	6	37.5	295	15	US-10-425-114-37694	Sequence 37694, A
37	6	37.5	319	15	US-10-282-122A-50380	Sequence 50380, A
38	6	37.5	330	14	US-10-155-435-10	Sequence 10, Appl
39	6	37.5	347	15	US-10-424-599-162707	Sequence 162707,
40	6	37.5	359	15	US-10-425-114-68412	Sequence 68412, A
41	6	37.5	367	17	US-10-425-115-337225	Sequence 337225,
42	6	37.5	396	17	US-10-425-115-276049	Sequence 276049,
43	6	37.5	413	17	US-10-425-115-287563	Sequence 287563,
44	6	37.5	421	15	US-10-374-780A-1522	Sequence 1522, Ap
45	6	37.5	421	15	US-10-412-699B-1593	Sequence 1593, Ap

ALIGNMENTS

RESULT 1
US-10-170-385-389
; Sequence 389, Application US/10170385
; Publication No. US20030203372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingeman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 532682000100
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 389
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-170-385-389

Query Match 50.0%; Score 8; DB 14; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YSLPKSEF 8

Db 145 YSLPKSEF 152
|||||

RESULT 2

US-10-425-115-220040
; Sequence 220040, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 220040
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(54)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: MRT4577_132263C.1.pep
US-10-425-115-220040

Query Match 43.8%; Score 7; DB 17; Length 54;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AVPDLEL 15
|||||
Db 10 AVPDLEL 16

RESULT 3

US-09-833-245-1222
; Sequence 1222, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1222
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-1222

Query Match 37.5%; Score 6; DB 11; Length 39;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 VPDLEL 15
|||||
Db 4 VPDLEL 9

RESULT 4

US-09-833-245-1224
; Sequence 1224, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1224
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-1224

Query Match 37.5%; Score 6; DB 11; Length 39;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 VPDLEL 15
|||||
Db 4 VPDLEL 9

RESULT 5

US-10-437-963-105909
; Sequence 105909, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, Thomas J.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 105909
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_10402C.1.pep
US-10-437-963-105909

Query Match 37.5%; Score 6; DB 16; Length 51;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKS 6
|||||
Db 10 YSLPKS 15

RESULT 6

US-10-425-115-252984
; Sequence 252984, Application US/10425115
; Publication No. US20040214272A1

; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 252984
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_162304C.1.pep
US-10-425-115-252984

Query Match 37.5%; Score 6; DB 17; Length 71;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 VPDLEL 15
Db 14 VPDLEL 19

RESULT 7
US-10-437-963-188997
; Sequence 188997, Application US/10437963
; Publication No. US20040123342A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 188997
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85548C.1.pep
US-10-437-963-188997

Query Match 37.5%; Score 6; DB 16; Length 86;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FAVPDL 13
Db 45 FAVPDL 50

RESULT 8
US-10-424-599-247004
; Sequence 247004, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 247004
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_65074C.1.pep
US-10-424-599-247004

Query Match 37.5%; Score 6; DB 15; Length 92;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 PDLELP 16
Db 43 PDLELP 48

RESULT 9
US-09-925-299-1494
; Sequence 1494, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1494
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (93)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (94)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (95)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1494

Query Match 37.5%; Score 6; DB 9; Length 95;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 PDLELP 16
Db 54 PDLELP 59

RESULT 10
US-09-925-299-1494
; Sequence 1494, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.

```
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL02
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05803
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1494
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (93)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (94)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (95)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1494

Query Match          37.5%; Score 6; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 PDLELP 16
Db 54 PDLELP 59

RESULT 11
US-10-282-122A-46052
; Sequence 46052, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46052
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-10-282-122A-46052

Query Match          37.5%; Score 6; DB 15; Length 101;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLPKSE 7
Db 13 SLPKSE 18

RESULT 12
US-10-424-599-188964
; Sequence 188964, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 188964
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_141649C.1.pep
US-10-424-599-188964

Query Match          37.5%; Score 6; DB 15; Length 116;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KSEFAV 10
Db 86 KSEFAV 91

RESULT 13
US-10-425-115-358214
; Sequence 358214, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 358214
; LENGTH: 117
```

```
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(117)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_89859C.1.pep
US-10-425-115-358214

Query Match          37.5%; Score 6; DB 17; Length 117;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 PDLELP 16
Db 18 PDLELP 23

RESULT 14
US-10-424-599-284187
; Sequence 284187, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 284187
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_98646C.1.pep
US-10-424-599-284187

Query Match          37.5%; Score 6; DB 15; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 PDLELP 16
Db 107 PDLELP 112

RESULT 15
US-10-437-963-190106
; Sequence 190106, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 190106
; LENGTH: 127
; TYPE: PRT
```

```
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_86551C.1.pep
US-10-437-963-190106

Query Match          37.5%; Score 6; DB 16; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FAVPDL 13
Db 110 FAVPDL 115

Search completed: November 17, 2004, 13:21:37
Job time : 4.8405 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 09:38:53 ; Search time 0.930043 Seconds
(without alignments)
1854.686 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 16

Sequence: 1 YSLPKSEFAVPDLELP 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	50.0	162	2	SI13195
2	8	50.0	193	2	I54178
3	8	50.0	200	2	S22411
4	7	43.8	193	2	S35613
5	7	43.8	369	2	F69146
6	7	43.8	409	2	B69147
7	6	37.5	133	2	S44579
8	6	37.5	136	2	AH3394
9	6	37.5	189	2	G72294
10	6	37.5	253	2	C98087
11	6	37.5	283	2	C70390
12	6	37.5	302	2	AB3573
13	6	37.5	320	2	F90738
14	6	37.5	320	2	H85568
15	6	37.5	320	2	H64816
16	6	37.5	329	2	T52579
17	6	37.5	379	2	T19588
18	6	37.5	413	2	T02463
19	6	37.5	455	2	T47921
20	6	37.5	490	2	F89933
21	6	37.5	491	2	C83206
22	6	37.5	503	2	S67390
23	6	37.5	668	2	S19739
24	6	37.5	715	2	T22787
25	6	37.5	802	2	G72720
26	6	37.5	985	2	T27083
27	6	37.5	1204	2	T19918
28	6	37.5	1225	2	T18954
29	6	37.5	1231	2	T24415

30 6 37.5 1369 2 T32338
31 6 37.5 1379 2 T13718
32 6 37.5 1690 2 T40847
33 6 37.5 2121 2 A59233
34 6 37.5 2124 2 A28452
35 6 37.5 2132 1 A55182
36 6 37.5 2453 2 S60254
37 6 37.5 3418 1 G02334
38 5 31.2 29 2 A61166
39 5 31.2 35 2 PS0439
40 5 31.2 35 2 PC4444
41 5 31.2 52 1 JX0241
42 5 31.2 52 1 RUDV
43 5 31.2 53 2 E75136
44 5 31.2 54 1 RUPF
45 5 31.2 63 2 D81307

ALIGNMENTS

RESULT 1
SI13195
ganglioside M2 activator protein - human
C:Species: Homo sapiens (man)
C>Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: SI13195
R:Fuerst, W.; Schubert, J.; Machleidt, W.; Meyer, H.E.; Sandhoff, K.
Eur. J. Biochem. 192, 709-714, 1990
A:Title: The complete amino-acid sequences of human ganglioside GM2 activator protein and
A:Reference number: SI13195; MUID:91006165; PMID:2209618
A:Accession: SI13195
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-162 <FUE>
A:Cross-references: UNIPROT:P17900

Query Match 50.0%; Score 8; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEF 8
Db 114 YSLPKSEF 121

RESULT 2
I54178
ganglioside M2 activator protein precursor - human
C:Species: Homo sapiens (man)
C>Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
C:Accession: I54178; JQ1037; S05036; S24410; S17107
R:Xie, B.; Kennedy, J.L.; McInnes, B.; Auger, D.; Mahuran, D.
Genomics 14, 796-798, 1992
A:Title: Identification of a processed pseudogene related to the functional gene encoding man chromosome 5.
A:Reference number: I54178; MUID:93052421; PMID:1427911
A:Accession: I54178
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-193 <RES>
A:Cross-references: UNIPROT:P17900; GB:L01439; NID:G183358; PIDN:AAA52767.1; PID:G183359
R:Xie, B.; McInnes, B.; Neote, K.; Lambornwah, A.M.; Mahuran, D.
Biochem. Biophys. Res. Commun. 177, 1217-1223, 1991
A:Title: Isolation and expression of a full-length cDNA encoding the human GM2 activator
A:Reference number: JQ1037; MUID:91282768; PMID:2059210
A:Accession: JQ1037
A:Molecule type: mRNA
A:Residues: 1-18, 'A', 20-193 <XIE>
A:Cross-references: GB:M76477; NID:G183356; PIDN:AAA35907.1; PID:G183357
A:Experimental source: HeLa cell
A:Note: 19-Thr and 69-Met were also found
R:Schoeder, M.; Klima, H.; Nakano, T.; Kwon, H.; Quintern, L.E.; Gaertner, S.; Suzuki, T.

FEBS Lett. 251, 197-200, 1989
A;Title: Isolation of a cDNA encoding the human G(M2) activator protein.
A;Reference number: S05036; MUID:89325664; PMID:2753159
A;Accession: S05036
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 15-193 <SCH>
A;Cross-references: GB:X16087; NID:g31852; PIDN:CAA34215.1; PID:g31853
R;Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A;Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A;Reference number: S22410; MUID:92207171; PMID:1554364
A;Accession: S22410
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 43-142, 'VST' <NAG>
A;Cross-references: EMBL:X61094; NID:g31854; PIDN:CAA43407.1; PID:g31855
C;Comment: This protein transports GM2 ganglioside from the lysosomal membrane to hexosa
C;Genetics:
A;Gene: GDB:GM2A
A;Cross-references: GDB:120000; OMIM:272750
F;1-23/Domain: signal sequence #status predicted <SIG>
F;32-193/Product: GM2 ganglioside activator protein #status predicted <GM2>
F;32-193/Product: ganglioside M2 activator #status predicted <MAT>
Query Match 50.0%; Score 8; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YSLPKSEF 8
DB 145 YSLPKSEF 152
RESULT 3
S22411
ganglioside M2 activator protein (clone pGAP2) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
A;Accession: S22411; S17108
R;Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A;Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A;Reference number: S22410; MUID:92207171; PMID:1554364
A;Accession: S22411
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-200 <NAG>
A;Cross-references: UNIPROT:P17900; EMBL:X61095; NID:g31856; PIDN:CAA43408.1; PID:g31857
Query Match 50.0%; Score 8; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YSLPKSEF 8
DB 152 YSLPKSEF 159
RESULT 4
S35613
ganglioside M2 activator protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 07-May-1999
A;Accession: S35613
R;Bellachio, G.; Stirling, J.L.; Orlicchio, A.; Beccari, T.
Biochem. J. 294, 227-230, 1993
A;Title: Cloning and sequence analysis of a cDNA clone coding for the mouse G(M2) activa
A;Reference number: S35613; MUID:93371367; PMID:7689829
A;Accession: S35613
A;Status: preliminary
A;Molecule type: mRNA

A;Residues: 1-193 <BEL>
Query Match 43.8%; Score 7; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 VPDLELP 16
DB 154 VPDLELP 160
RESULT 5
F69146
teichoic acid biosynthesis protein RodC related protein - Methanobacterium thermoautotot
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
A;Accession: F69146
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: F69146
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-369 <MTH>
A;Cross-references: UNIPROT:O26461; GB:AE000822; GB:AE000665; NID:g2621420; PIDN:AAB8486
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH361
Query Match 43.8%; Score 7; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 FAVPDLE 14
DB 291 FAVPDLE 297
RESULT 6
B69147
teichoic acid biosynthesis protein RodC related protein - Methanobacterium thermoautotot
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
A;Accession: B69147
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: B69147
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-409 <MTH>
A;Cross-references: UNIPROT:O26465; GB:AE000822; GB:AE000666; NID:g2621420; PIDN:AAB8487
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH365
Query Match 43.8%; Score 7; DB 2; Length 409;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 FAVPDLE 14
DB 332 FAVPDLE 338
RESULT 7
S44579

hypothetical protein YKL603 - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 09-Jul-2004
C:Accession: S44579
R:Vandenbol, M.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.
Yeast 10, 35-40, 1994
A>Title: DNA sequencing of a 36.2 kb fragment located between the FAS1 and LAP4 loci of
A:Reference number: S44563
A:Accession: S44579
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-133 <VAN>
A:Cross-references: UNIPROT:Q05653; EMBL:Z26877; NID:G407482; PIDN:CAA81505.1; PID:G4074
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1993
C:Superfamily: Saccharomyces hypothetical protein YKL603

Query Match 37.5%; Score 6; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLPKSE 7
|||||
Db 49 SLPKSE 54

RESULT 8
AH3394
lactoylglutathione lyase (EC 4.4.1.5) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AH3394
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AH3394
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-136 <KUR>
A:Cross-references: UNIPROT:Q8VGL6; GB:AE008917; PIDN:AAJ52323.1; PID:g17983117; GSPDB:G
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI1142
A:Map position: I
C:Superfamily: hypothetical protein AF2218
C:Keywords: carbon-sulfur lyase

Query Match 37.5%; Score 6; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AVPDLE 14
|||||
Db 13 AVPDLE 18

RESULT 9
G72294
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: G72294
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.

Nature 399, 323-329, 1999
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; PMID:99287316; PMID:10360571
A:Accession: G72294
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <ARN>

A:Cross-references: UNIPROT:Q9X015; GB:AE001769; GB:AE0000512; NID:G4981639; PIDN:AAD36177;
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TMI101
C:Superfamily: VPS29-like phosphoesterase-related protein

Query Match 37.5%; Score 6; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLPKSE 7
|||||
Db 150 SLPKSE 155

RESULT 10
C98087
oxidoreductase (EC 1.-.-.-) [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: C98087
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Be
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Mc
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; PMID:21429245; PMID:11544234
A:Accession: C98087
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-253 <KUR>
A:Cross-references: UNIPROT:Q8DNH3; GB:AE007317; PIDN:AAJ00528.1; PID:g15459404; GSPDB:G
C:Genetics:
A:Gene: ydfg
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase

Query Match 37.5%; Score 6; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KSEFAV 10
|||||
Db 43 KSEFAV 48

RESULT 11
C70390
formate dehydrogenase formation protein FdhE - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: C70390
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A>Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; PMID:98196666; PMID:9537320
A:Accession: C70390
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-283 <AQF>
A:Cross-references: UNIPROT:O67150; GB:AE000720; NID:g2983529; PIDN:AA07106.1; PID:g2983
A:Experimental source: strain VF5
C:Genetics:
A:Gene: fdhE
C:Superfamily: formate dehydrogenase accessory protein FdhE

Query Match 37.5%; Score 6; DB 2; Length 283;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AVPDLE 14

```

Db          252 AVPDLE 257
|||||
RESULT 12
AB3573
oligopeptide transport ATP-binding protein oppD BMEII0507 [imported] - Brucella melitensis
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AB3573
R:DelVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AB3573
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <KUR>
A:Cross-references: UNIPROT:Q8YXK6; GB:AE008918; PIDN:AAL53749.1; PID:gl7984675; GSPDB:G
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEII0507
A:Map position: 11
C:Superfamily: inner membrane protein malK; ATP-binding cassette homology
Query Match          37.5%; Score 6; DB 2; Length 302;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY          9 AVPDLE 14
|||||
Db          293 AVPDLE 298
|||||
RESULT 13
F90738
probable enzyme [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: F90738
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Iehii, K.; Yokoyama, K.; Han, C.G.
Gasawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90738
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-320 <HAY>
A:Cross-references: UNIPROT:Q8YXK6; GB:BA000007; PIDN:BA834301.1; PID:gl3360337; GSPDB:G
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECG878
Query Match          37.5%; Score 6; DB 2; Length 320;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY          10 VPDLEL 15
|||||
Db          37 VPDLEL 42
|||||
RESULT 14
H85588
probable enzyme ybiB [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: H85588
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001

```

```

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85588
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-320 <STO>
A:Cross-references: UNIPROT:Q8YXK6; GB:AE005174; NID:gl2513796; PIDN:AAG55172.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ybiB
Query Match          37.5%; Score 6; DB 2; Length 320;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY          10 VPDLEL 15
|||||
Db          37 VPDLEL 42
|||||

```

RESULT 15

```

H64816
ybiB protein - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: H64816
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64816
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-320 <BLAT>
A:Cross-references: UNIPROT:P30177; GB:AE000182; GB:U00096; NID:gl787015; PIDN:AAC73887
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ybiB
C:Keywords: transmembrane protein
F:98-114/Domain: transmembrane #status predicted <TM>
Query Match          37.5%; Score 6; DB 2; Length 320;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY          10 VPDLEL 15
|||||
Db          37 VPDLEL 42
|||||

```

```

Search completed: November 17, 2004, 10:47:53
Job time : 2.83004 secs

```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 09:35:07 ; Search time 4.63337 Seconds
(without alignments)
1986.889 Million cell updates/sec

Title: US-10-030-937-72
Perfect score: 16
Sequence: 1 YSLPKSEFAVPDLLELP 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : UniProt 02:.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	50.0	189	2	Q6LBL5 homo sapien
2	8	50.0	189	2	Caa43994
3	8	50.0	190	2	Q8HXX6
4	8	50.0	193	1	SAP3 HUMAN
5	7	43.8	193	1	SAP3 MOUSE
6	7	43.8	199	2	O6IN37
7	7	43.8	199	2	O8CJH4
8	7	43.8	369	2	O26461
9	7	43.8	409	2	O26465
10	7	43.8	449	2	O15717
11	7	43.8	523	2	Q7NSP6
12	6	37.5	119	2	Q9LVP5
13	6	37.5	121	2	Q85624
14	6	37.5	126	2	O8IGV6
15	6	37.5	126	2	Q8MWD0
16	6	37.5	126	2	AAAT0795
17	6	37.5	133	2	Q05653
18	6	37.5	134	2	Q8G1A1
19	6	37.5	136	2	Q8YGL6
20	6	37.5	151	2	Q8NR47
21	6	37.5	151	2	CAF19921
22	6	37.5	164	1	FBP2_EPIPO
23	6	37.5	173	2	Q98IT5
24	6	37.5	174	2	Q6DIZ9
25	6	37.5	175	2	Q8P375
26	6	37.5	176	1	MCEE HUMAN
27	6	37.5	176	2	Q8JSZ0
28	6	37.5	178	1	MCEE MOUSE
29	6	37.5	181	2	O6Y2G4
30	6	37.5	181	2	O6Y2G7
31	6	37.5	181	2	O6Y2I2

32 6 37.5 181 2 AAP02897
33 6 37.5 181 2 AAP02900
34 6 37.5 181 2 AAP02914
35 6 37.5 189 2 Q9X0I5
36 6 37.5 202 2 Q97CD4
37 6 37.5 208 2 Q722Y5
38 6 37.5 208 2 AAT03376
39 6 37.5 214 1 TESC HUMAN
40 6 37.5 214 1 TESC MOUSE
41 6 37.5 214 2 Q6D7Z5
42 6 37.5 214 2 Q8VCN1
43 6 37.5 216 2 Q9NWT9
44 6 37.5 220 2 Q8GVG9
45 6 37.5 227 2 O07837

ALIGNMENTS

RESULT 1
Q6LBL5
ID Q6LBL5 PRELIMINARY; PRT; 189 AA.
AC Q6LBL5;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE GM2 activator protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
human GM2 activator protein.";
RL FEBS Lett. 289:260-264 (1991).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9327527; PubMed=8503891;
RA Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA Sandhoff K.;
RT "Over-expression of a functionally active human GM2-activator protein
in escherichia coli.";
RL Biochem. J. 292:571-576 (1993).
DR EMBL: X62078; CAA43994.1; -;
DR InterPro: IPR003172; EI_DerP2_DerF2.
DR SMART: SM00737; ML; 1.
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;

Query Match 50.0%; Score 8; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEF 8
|||||||
Db 141 YSLPKSEF 148

RESULT 2
CAA43994
ID CAA43994 PRELIMINARY; PRT; 189 AA.
AC CAA43994;
DT 02-MAR-2004 (Tremblrel. 27, Created)
DT 02-MAR-2004 (Tremblrel. 27, Last sequence update)
DT 02-MAR-2004 (Tremblrel. 27, Last annotation update)
DE GM2 activator protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

RN  SEQUENCE FROM N.A.
RP  MEDLINE=92008637; PubMed=1915857;
RX  Kikina H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA  Sandhoff K.;
RT  "Characterization of full-length cDNAs and the gene coding for the
RT  human GM2 activator protein.";
RL  FEBS Lett. 289:260-264(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93277527; PubMed=8503891;
RA  Kikina H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
RA  Sandhoff K.;
RT  "Over-expression of a functionally active human GM2-activator protein
RT  in Escherichia coli.";
RL  Biochem. J. 292:571-576(1993).
DR  EMBL; X62078; CA443994.1; -
SQ  SEQUENCE 189 AA, 20362 MW, 9B8C7F18DC7439BE CRC64;

Query Match      50.0%; Score 8; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 YSLPKSEF 8
Db  141 YSLPKSEF 148

RESULT 3
QBHX6 Q8HXX6 PRELIMINARY; PRT; 190 AA.
AC Q8HXX6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Ganglioside GM2 activator.
GN Name=gma2;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN  [1]
RP  SEQUENCE FROM N.A.
RC TISSUE=Brain cerebellum cortex;
RA Kusuda J., Osada N., Hida M., Sugano S., Hashimoto K.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB083313; BAC20592.1; -
DR HSSP; P17900; IG13.
DR InterPro; IPR003172; E1_DerP2_DerF2.
DR SMART; SM00737; MU; 1.
SQ  SEQUENCE 190 AA, 20494 MW, 9F9582BEB75715C3 CRC64;

Query Match      50.0%; Score 8; DB 2; Length 190;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 YSLPKSEF 8
Db  142 YSLPKSEF 149

RESULT 4
SAP3 HUMAN STANDARD; PRT; 193 AA.
ID SAP3 HUMAN
AC P17900; Q14426; Q14428;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate
DE activator protein) (Shingolipid activator protein 3) (SAP-3).
GN Name=GM2A;
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN  [1]
RP  SEQUENCE FROM N.A., AND VARIANT ALA-19.
RX  MEDLINE=91282768; PubMed=2059210;
RA  Xie B., McInnes B., Neote K., Lamhonwah A.-M., Mahuran D.;
RT  "Isolation and expression of a full-length cDNA encoding the human G-
RT  M2 activator protein.";
RL  Biochem. Biophys. Res. Commun. 177:1217-1223(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=92008637; PubMed=1915857;
RA  Kikina H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
RA  Sandhoff K.;
RT  "Characterization of full-length cDNAs and the gene coding for the
RT  human GM2 activator protein.";
RL  FEBS Lett. 289:260-264(1991).
RN  [3]
RP  SEQUENCE FROM N.A., AND VARIANT ALA-19.
RC TISSUE=Placenta;
RX  MEDLINE=92207171; PubMed=1554364;
RA  Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
RT  "Evidence for two cDNAs encoding human GM2-activator protein.";
RL  Biochem. J. 282:807-813(1992).
RN  [4]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93052421; PubMed=1427911;
RA  Xie B., Kennedy J.L., McInnes B., Auger D., Mahuran D.J.;
RT  "Identification of a processed pseudogene related to the functional
RT  gene encoding the GM2 activator protein: localization of the
RT  pseudogene to human chromosome 3 and the functional gene to human
RT  chromosome 5.";
RL  Chromosoma 14:796-798(1992).
RN  [5]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=99294584; PubMed=10364519;
RA  Chen B., Rigat B., Curry C., Mahuran D.J.;
RT  "Structure of the GM2A gene: identification of an exon 2 nonsense
RT  mutation and a naturally occurring transcript with an in-frame
RT  deletion of exon 2.";
RL  Am. J. Hum. Genet. 65:77-87(1999).
RN  [6]
RP  SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX  MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA  Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA  Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA  Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA  Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA  Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA  Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA  Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA  Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA  Rosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA  Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA  Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA  Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA  Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA  Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA  Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA  Butterfield Y.S.N., Krzywicki M.J., Skalska U., Smalish D.E.,
RA  Schnerch A., Schein J.E., Jones S.O.M., Marra M.A.;
RT  "Generation and initial analysis of more than 15,000 full-length human
RT  and mouse cDNA sequences.";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN  [7]
RP  SEQUENCE OF 15-193 FROM N.A.
RX  MEDLINE=89325664; PubMed=2753159;
RA  Schroeder M., Kikina H., Nakano T., Kwon H., Quintern L.E.,
RA  Gaertner S., Suzuki K., Sandhoff K.;
RT  "Isolation of a cDNA encoding the human GM2 activator protein.";
RL  FEBS Lett. 251:197-200(1989).

```

(8)
 RP SEQUENCE OF 32-193.
 RC TISSUE-Kidney;
 RX MEDLINE=91006165; PubMed=2209618;
 RA Furst W., Schubert J., Machleidt W., Meyer H.E., Sandhoff K.;
 RT "The complete amino-acid sequences of human ganglioside GM2 activator
 RL protein and cerebroside sulfate activator protein.";
 RL Eur. J. Biochem. 192:709-714(1990).
 [9]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=20545600; PubMed=11090283; DOI=10.1006/jmbi.2000.4225;
 RA Wright C.S., Li S.-C., Rastinejad F.;
 RT "Crystal structure of human GM2-activator protein with a novel beta-
 RL cup topology.";
 RL J. Mol. Biol. 304:411-422(2000).
 [10]
 RP VARIANT TSD-AB ARG-138.
 RX MEDLINE=92008638; PubMed=1915858;
 RA Schroeder M., Schnabel D., Suzuki K., Sandhoff K.;
 RT "A mutation in the gene of a glycolipid-binding protein (GM2
 RL activator) that causes GM2-gangliosidosis variant AB.";
 RL FEBS Lett. 290:1-3(1991).
 [11]
 RP VARIANT TSD-AB PRO-169.
 RX MEDLINE=94063850; PubMed=8244332;
 RA Schroeder M., Schnabel D., Hurwitz R., Young E., Suzuki K.,
 RT Sandhoff K.;
 RT "Molecular genetics of GM2-gangliosidosis AB variant: a novel mutation
 RL and expression in BHK cells.";
 RL Hum. Genet. 92:437-440(1993).
 [12]
 RP VARIANT TSD-AB LYS-88 DEL.
 RX MEDLINE=97055887; PubMed=8900233;
 RA Schepers U., Glombitza G., Lemm T., Hoffmann A., Chabas A., Ozand P.,
 RT Sandhoff K.;
 RT "Molecular analysis of a GM2-activator deficiency in two patients with
 RL GM2-gangliosidosis AB variant.";
 RL Am. J. Hum. Genet. 59:1048-1056(1996).
 CC -I- FUNCTION: Binds gangliosides and stimulates ganglioside GM2
 CC degradation. It stimulates only the breakdown of ganglioside GM2
 CC and glycolipid GA2 by beta-hexosaminidase A. It extracts single
 CC GM2 molecules from membranes and presents them in soluble form to
 CC beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and
 CC conversion to GM3.
 CC -I- SUBCELLULAR LOCATION: Lysosomal.
 CC -I- DISEASE: Defects in GM2A are the cause of Tay-Sachs disease AB
 CC variant (TSD-AB) [MIM:272750]; also known as GM2-gangliosidosis
 CC type AB.
 CC WWW="http://www.hexdb.mcgill.ca/?Topic=GM2A&Page=MutationSubmission".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M76477; AAA35907.1; -;
 CC EMBL; X62078; CAA43993.1; -;
 CC EMBL; X61095; CAA43408.1; ALT_INIT.
 CC EMBL; L01439; AAA52767.1; -;
 CC EMBL; AF124719; AAD25741.1; -;
 CC EMBL; AF124717; AAD25741.1; JOINED.
 CC EMBL; AF124718; AAD25741.1; JOINED.
 CC EMBL; BC009273; AAH09273.1; -;
 CC EMBL; X16087; CAA34215.1; -;
 CC PIR; I54178; I54178.
 CC PIR; S13195; S13195.
 CC PIR; S22411; S22411.
 CC PDB; 1G13; X-ray; A/B/C=32-193.
 CC Genew; HGNC:4367; GM2A.

DR MIM; 272750; -.
 DR GO; GO:0005764; C:lysosome; NAS.
 DR GO; GO:0030230; F:sphingolipid activator protein activity; NAS.
 DR GO; GO:0019377; P:glycolipid catabolism; NAS.
 DR GO; GO:0030149; P:sphingolipid catabolism; NAS.
 DR InterPro; IPR003172; EI_DerP2_DerF2.
 DR SMART; SM00737; ML; 1.
 KW 3D-structure; Direct protein sequencing; Disease mutation;
 KW Glycoprotein; GM2-gangliosidosis; Lysosome; Polymorphism; Signal;
 KW Sphingolipid metabolism.
 FT SIGNAL 1 31 Ganglioside GM2 activator.
 FT CHAIN 32 193
 FT DISULFID 39 183
 FT DISULFID 99 106
 FT DISULFID 112 138
 FT DISULFID 125 136
 FT CARBOHYD 63 63
 FT VARIANT 19 19 N-linked (GlcNAc...).
 FT T -> A.
 FT /FTID=VAR_013830.
 FT Missing (in 80% of the protein).
 FT /FTID=VAR_006946.
 FT Missing (in TSD-AB).
 FT /FTID=VAR_011697.
 FT C -> R (in TSD-AB).
 FT /FTID=VAR_006947.
 FT R -> P (in TSD-AB).
 FT /FTID=VAR_011698.
 FT V -> I (in Ref. 3).
 FT V -> M (in Ref. 3).
 FT CONFLICT 59 59
 FT CONFLICT 69 69
 FT STRAND 35 38
 FT TURN 41 43
 FT STRAND 46 54
 FT STRAND 58 59
 FT STRAND 63 72
 FT STRAND 76 76
 FT STRAND 81 90
 FT TURN 91 92
 FT STRAND 93 96
 FT STRAND 100 100
 FT TURN 101 102
 FT STRAND 103 103
 FT STRAND 107 108
 FT TURN 109 110
 FT HELIX 111 118
 FT TURN 121 122
 Query Match 50.0%; Score 8; DB 1; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YSLPKSEF 8
 DB 145 YSLPKSEF 152
 RESULT 5
 SAP3 MOUSE STANDARD; PRT; 193 AA.
 AC Q60618; Q61610; Q61819;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate
 DE activator protein) (Shingolipid activator protein 3) (SAP-3).
 GN Name=Gm2a;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
 RX MEDLINE=95229165; PubMed=7713516;

FT	CHAIN	32	193	Ganglioside GM2 activator
FT	DISULFID	39	183	By similarity.
FT	DISULFID	99	106	By similarity.
FT	DISULFID	112	138	By similarity.
FT	DISULFID	125	136	By similarity.
FT	CARBOHYD	151	131	N-linked (GlcNac..) (Potential).
FT	CONFLICT	53	53	I -> T (in Ref. 1).
SQ	SEQUENCE	193 AA;	20824 MW;	59C4C4ABE56FA1FC7 CRC64;

Query Match 43.8%; Score 7; DB 1; Length 193;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	10	VPDLELP	16	
Db	154	VPDLELP	160	

RESULT 6	ID	Q6IN37	PRELIMINARY;	PRT;	199 AA.
Q6IN37	AC	Q6IN37;			
AC	DT	05-JUL-2004 (TrEMBLrel. 27, Created)			
DT	DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)			
DT	DT	05-JUL-2004 (TrEMBLrel. 27, Last annotation update)			
DE	DE	GM2 ganglioside activator protein.			
DE	GN	Name=Gm2a;			
OS	OS	Rattus norvegicus (Rat).			
OC	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OC	NCBI_TaxID=10116;				
RN	RN	[1]			
RP	RP	SEQUENCE FROM N.A.			
RP	RC	TISUE=Lung;			
RX	RX	MEDLINE=22388257; PubMed=12477932;			
RA	RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	RA	Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,			
RA	RA	Diatchenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	RA	Stapleton M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	RA	Brownstein M.J., Gaudet J., Brownstein R., Robinson J., Mullen J.,			
RA	RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,			
RA	RA	Bosak S.A., McWray P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	RA	Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,			
RA	RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	RA	Blatesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,			
RA	RA	Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,			
RA	RA	Jones S.J., Marra M.A.;			
RT	RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	RT	and mouse cDNA sequences.";			
RL	RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RL	RL	[2]			
RP	RP	SEQUENCE FROM N.A.			
RP	RC	TISUE=Lung;			
RA	RA	Strausberg R.;			
RL	RL	Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.			
DR	DR	EMBL; BC072474; AAH72474.1; -.			
DR	DR	InterPro; IPR003172; B1_DerP2_DerF2.			
DR	DR	SMART; SM00737; ML; 1			
SQ	SEQUENCE	199 AA;	21493 MW;	C2B5203DFDFABF507 CRC64;	

Query Match 43.8%; Score 7; DB 2; Length 199;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	10	VPDLELP	16	
Db	160	VPDLELP	166	

```

RESULT 7
Q8CJH4
ID Q8CJH4 PRELIMINARY; PRT; 199 AA.
AC Q8CJH4;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE GW2 activator protein.
GN Name=IGMZAP;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Miwa N., Okada T., Nakamura S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051391; BAC24018.1; -.
DR HSSP; P17900; IG13.
DR InterPro; IPR003172; EL_DerP2_DerF2.
DR SMART; SMO0737; ML; 1.
SQ SEQUENCE 199 AA; 21521 MW; D585203FDFABF507 CRC64;

Query Match 43.8%; Score 7; DB 2; Length 199;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 VPDLELP 16
Db 160 VPDLELP 166

RESULT 8
O26461
ID O26461 PRELIMINARY; PRT; 369 AA.
AC O26461;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Teichoic acid biosynthesis protein RodC related protein.
GN OrderedLocusNames=MTH361;
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT delah: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000822; AAB84867.1; -.
DR PIR; F69146; F69146.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0047355; F:CDP-glycerol glycerophosphotransferase acti. .; IEA.
DR GO; GO:0019350; P:teichoic acid biosynthesis; IEA.
DR InterPro; IPR007554; Glyphos transf.
DR Pfam; PF04464; Glyphos_transf; 1.
KW Complete proteome.
SQ SEQUENCE 369 AA; 43283 MW; 89CFCEBA13562A8C CRC64;

Query Match 43.8%; Score 7; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 VPDLELP 16
Db 160 VPDLELP 166

RESULT 9
O26465
ID O26465 PRELIMINARY; PRT; 409 AA.
AC O26465;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Teichoic acid biosynthesis protein RodC related protein.
GN OrderedLocusNames=MTH365;
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT delah: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000822; AAB84871.1; -.
DR PIR; B69147; B69147.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0047355; F:CDP-glycerol glycerophosphotransferase acti. .; IEA.
DR GO; GO:0019350; P:teichoic acid biosynthesis; IEA.
DR InterPro; IPR007554; Glyphos transf.
DR Pfam; PF04464; Glyphos_transf; 1.
KW Complete proteome.
SQ SEQUENCE 409 AA; 47677 MW; DFFED3ED6F65155F CRC64;

Query Match 43.8%; Score 7; DB 2; Length 409;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FAVPDLE 14
Db 332 FAVPDLE 338

RESULT 10
O15717
ID O15717 PRELIMINARY; PRT; 449 AA.
AC O15717;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE PotA (Fragment).
GN Name=potA;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA Shaulsky G., Loomis W.F.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF015567; AAB69391.1; -.
DR DictyBase; DDB0214952; potA.
FT NON_TER 1 1
FT NON_TER 449 449

```

```
SQ SEQUENCE 449 AA; 50687 MW; 0F1C85E03996E2DE CRC64;
Query Match 43.8%; Score 7; DB 2; Length 449;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AVPDLEL 15
|||||
Db 249 AVPDLEL 255

RESULT 11
Q7NSP6 PRELIMINARY; PRT; 523 AA.
AC Q7NSP6;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Similar to unknown protein of Photorhabdus.
GN OrderedLocNames=plu1900;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Derose R., Derzelle S., Freyssinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Signier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RA "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens.";
RL Nat. Biotechnol. 21:1307-1313(2003).
DR EMBL; BX571865; CAE14193.1; -.
DR PhotoList; plu1900; -.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
KW Complete proteome.
SQ SEQUENCE 523 AA; 56514 MW; C636009FCC927814 CRC64;

Query Match 43.8%; Score 7; DB 2; Length 523;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSE 7
|||||
Db 28 YSLPKSE 34

RESULT 12
Q9LVP5 PRELIMINARY; PRT; 119 AA.
AC Q9LVP5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MXE2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
```

```
RL DNA Res. 7:131-135(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB018121; BAB01990.1; -.
SQ SEQUENCE 119 AA; 13680 MW; 52AF9961891CCF82 CRC64;

Query Match 37.5%; Score 6; DB 2; Length 119;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 PDLELP 16
|||||
Db 113 PDLELP 118

RESULT 13
Q856Z4 PRELIMINARY; PRT; 121 AA.
AC Q856Z4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Gp40.
OS Mycobacteriophage Rosebush.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=205874;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22592660; PubMed=12705866;
RA Pedulla M.L., Ford M.E., Houtz J.M., Karthikeyan T., Wadsworth C.,
RA Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Panunzio N.R.,
RA Brucker W., Kumar V., Lawrence J.G., Jacobs W.J., Keenan L., Bardarov S.,
RA Kriakov J., Lawrence J.G., Hendrix R.W.,
RA Hatfull G.F.;
RA "Origins of highly mosaic mycobacteriophage genomes.";
RL Cell 113:171-182(2003).
DR EMBL; AY129334; AAN01882.1; -.
SQ SEQUENCE 121 AA; 13251 MW; 9B4E586066341C35 CRC64;

Query Match 37.5%; Score 6; DB 2; Length 121;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SEFAVP 11
|||||
Db 99 SEFAVP 104

RESULT 14
Q8IGJ6 PRELIMINARY; PRT; 126 AA.
AC Q8IGJ6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE RH04491P.
GN ORFNames=CG30423;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celnikner S.;
```

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BT001754; AAN71509.1; -

DR FlyBase; FBgn0050423; CG30423.

DR InterPro; IPR007262; Vps55.

DR Pfam; PF04133; Vps55; 1.

SQ SEQUENCE 126 AA; 14013 MW; 7D159BED8507B729 CRC64;

Query Match

Best Local Similarity 37.5%; Score 6; DB 2; Length 126;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PKSEFA 9

Db 63 PKSEFA 68

RESULT 15

Q8MMD0

ID Q8MMD0 PRELIMINARY; PRT; 126 AA.

AC Q8MMD0;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE CG30423-PB.

GN ORFNames=CG30423;

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=72027;

RP SEQUENCE FROM N.A.

RP MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,

RA Yeh R.F., Zaveri J.S., Zhang F.N., Zhong W., Zhou X., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195 (2000).

RN SEQUENCE FROM N.A.

RX MEDLINE=22426065; PubMed=12537568;

RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,

RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,

RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,

RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,

RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,

RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;

RT "Finishing a whole-genome shotgun: release 3 of the Drosophila

melanogaster euchromatic genome sequence."

RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).

RN [3]

RX SEQUENCE FROM N.A.

RP MEDLINE=22426070; PubMed=12537573;

RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,

RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,

RA Ashburner M., Celniker S.E.;

RT "The transposable elements of the Drosophila melanogaster euchromatin:

a genomic perspective."

RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).

RN [4]

RX SEQUENCE FROM N.A.

RP MEDLINE=22426069; PubMed=12537572;

RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,

RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

RA Lewis S.E.;

RT "Annotation of the Drosophila melanogaster euchromatic genome: a

systematic review."

RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).

RN [5]

RX SEQUENCE FROM N.A.

RP FLYBASE;

RT Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RN [6]

RX SEQUENCE FROM N.A.

RP FLYBASE;

RT Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.

RL EMBL; AE003465; AAM70795.2; -

DR FlyBase; FBgn0050423; CG30423.

DR InterPro; IPR007262; Vps55.

DR Pfam; PF04133; Vps55; 1.

DR SQ SEQUENCE 126 AA; 13999 MW; 7EB59DE8507B129 CRC64;

Query Match 37.5%; Score 6; DB 2; Length 126;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PKSEFA 9

Db 63 PKSEFA 68

Search completed: November 17, 2004, 10:45:36
Job time : 6.63337 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 17:56:16 : Search time 651.378 Seconds
(without alignments)
1161.592 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 83

Sequence: 1 YSLPKSEFAVPDLPLP 16

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10030937/runat_16112004_153013_2913/app_query.fasta_1.789
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937@cgn_1_1_7771@runat_16112004_153013_2913 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOB=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	95.2	821	9 HSGM2AP	X16087 Human mRNA
2	79	95.2	953	9 HUMGM2	M76477 Human G-M2
3	79	95.2	1043	9 HUMGM2A	L01439 Human GM2-a
4	79	95.2	1045	6 CQ728078	CQ728078 Sequence

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	79	95.2	1093	9	HSGM2APB
6	79	95.2	1706	9	HSGM2A3
7	79	95.2	2413	9	BC009273
8	79	95.2	2436	6	AX330938
9	79	95.2	2436	9	HSGM2APT
10	79	95.2	107320	2	AC011391
c 11	79	95.2	120584	9	AC011342
c 12	79	95.2	151712	9	AC008385
13	72	86.7	950	9	AB083313
14	69	83.1	600	10	AB051391
15	69	83.1	1983	6	E12286
16	69	83.1	1983	6	AX827433
17	69	83.1	2028	10	BC072474
c 18	69	83.1	189516	2	AC128065
19	69	83.1	242756	2	AC093965
c 20	69	83.1	249474	2	AC136421
21	68	81.9	726	10	MMGM2APA
22	68	81.9	928	5	EX950406
23	68	81.9	1113	10	MUSGM2ACT
24	68	81.9	2003	10	MMU09816
25	68	81.9	2024	10	BC004651
26	68	81.9	39262	10	AL772357
27	66	79.5	529	9	AB051291
28	66	79.5	564	9	HUMGM2APC
29	66	79.5	111861	9	AC069435
30	60	72.3	470	4	AY553654
31	60	72.3	474	4	AY553653
c 32	55	66.3	186966	2	AC144390
c 33	55	66.3	239537	2	AC128804
34	54	65.1	100575	9	HSDJ81F6
35	51	61.4	198821	10	AL929545
36	51	61.4	220810	2	AC107735
37	51	61.4	22368	2	AC103652
38	50	60.2	145496	8	AC144342
39	50	60.2	216805	2	AC121058
c 40	50	60.2	238839	2	AC128615
41	50	60.2	246814	2	AC103482
42	50	60.2	248403	2	AC096264
43	49	59.0	3129	8	AK066228
44	49	59.0	3182	8	AF459087
45	49	59.0	8353	8	AF459086

ALIGNMENTS

RESULT 1	HSGM2AP	Human mRNA for G(M2) activator protein.	821 bp	mRNA	linear	PRI 19-JUL-1995
HSGM2AP	X16087					
LOCUS	X16087.1	GI:31852				
DEFINITION	G(M2) activator protein; G(M2) gangliosidosis.					
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
FEATURES						

Over-expression of a functionally active human GM2-activator protein in *Escherichia coli*
Biochem. J. 292 (Pt 2), 571-576 (1993)
8503891
Data kindly reviewed (23-JAN-1991) by Sandhoff K.
Location/Qualifiers

ORIGIN

Alignment Scores:

1. .821

Length: 953

Score: 79.00

Matches: 15

Percent Similarity: 93.75%

Conservative: 0

Best Local Similarity: 93.75%

Mismatches: 1

Query Match: 95.18%

Indels: 0

Gaps: 0

DB:

US-10-030-937-72 (1-16) x HUMGM2 (1-953)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16

Db 523 TACTCACTGCCCAAGAGCGAATTCGTTGCTGACCTGGAGCTGCC 570

RESULT 3

HUMGM2A

LOCUS

DEFINITION

Human GM2-activator protein (GM2A) mRNA, complete cds.

ACCESSION

L01439

VERSION

L01439.1

KEYWORDS

GM2 activator protein.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1043)

AUTHORS

Xie,B., Kennedy,J.L., McInnes,B., Auger,D. and Mahuran,D.

TITLE

Identification of a processed pseudogene related to the functional gene encoding the GM2 activator protein: localization of the pseudogene to human chromosome 3 and the functional gene to human chromosome 5

JOURNAL

Genomics 14 (3), 796-798 (1992)

MEDLINE

93052421

PUBMED

1427911

COMMENT

Original source text: Homo sapiens cDNA to mRNA.

FEATURES

Location/Qualifiers

1..1043

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/map="5"

1..1043

/gene="GM2A"

91..672

/gene="GM2A"

/codon_start=1

/product="GM2-activator protein"

/protein_id="AAA52767.1"

/db_xref="GI:183359"

/db_xref="GDB:G00-120-000"

/translation="MOSLMQAPLLIALGLLLATPAQAHLKPKSLSFSDNCDEGKDPAVIRSLTLEPDIIVPGNVLTSVWGTSVPLSPKVDLVEKEVAGLWKIPCTDYIGSCTFHFCDVLDMLIPTGCEPCEPLRTYGLPCHCFKPGEGTSLPKSEFVFPDLELP

ORIGIN

Alignment Scores:

1. .821

Length: 1043

Score: 79.00

Matches: 15

Percent Similarity: 93.75%

Conservative: 0

Best Local Similarity: 93.75%

Mismatches: 1

Query Match: 95.18%

Indels: 0

Gaps: 0

DB:

US-10-030-937-72 (1-16) x HUMGM2A (1-1043)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16

Db 523 TACTCACTGCCCAAGAGCGAATTCGTTGCTGACCTGGAGCTGCC 570

RESULT 4

source

1. .821

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="pGAP1"

/cell_type="fibroblast"

<1..540

/codon_start=1

/product="G(M2) activator protein"

/protein_id="CAA34215.1"

/db_xref="GI:31853"

/db_xref="GOA:PI7900"

/db_xref="Swiss-Prot:PI7900"

/translation="ILLATPAQAHLKPKSLSFSDNCDEGKDPVIRSLTLEPDIIVPGNVLTSVWGTSVPLSPKVDLVEKEVAGLWKIPCTDYIGSCTFHFCDVLDMLIPTGCEPCEPLRTYGLPCHCFKPGEGTSLPKSEFVFPDLELP

SSSKRLGICIKIAASLKGI"

<1..51

sig_peptide

mat_peptide

52..537

/product="G(M2) activator protein"

/evidence=experimental

ORIGIN

Alignment Scores:

2.23e-05

Length: 821

Score: 79.00

Matches: 15

Percent Similarity: 93.75%

Conservative: 0

Best Local Similarity: 93.75%

Mismatches: 1

Query Match: 95.18%

Indels: 0

Gaps: 0

DB:

US-10-030-937-72 (1-16) x HSGM2AP (1-821)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16

Db 391 TACTCACTGCCCAAGAGCGAATTCGTTGCTGACCTGGAGCTGCC 438

RESULT 2

HUMGM2

LOCUS

DEFINITION

Human G-M2 activator protein mRNA, complete cds.

ACCESSION

M76477.1

VERSION

M76477.1

KEYWORDS

G-M2 activator protein.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 953)

AUTHORS

Xie,B., McInnes,B., Neote,K., Lambonwah,A.M. and Mahuran,D.

TITLE

Isolation and expression of a full-length cDNA encoding the human GM2 activator protein

JOURNAL

Biochem. Biophys. Res. Commun. 177 (3), 1217-1223 (1991)

MEDLINE

91282768

PUBMED

2059210

COMMENT

Original source text: Homo sapiens mRNA.

FEATURES

Location/Qualifiers

1..953

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

91..672

/codon_start=1

/product="G-M2 activator protein"

/protein_id="AAA35907.1"

/db_xref="GI:183357"

/translation="MOSLMQAPLLIALGLLLATPAQAHLKPKSLSFSDNCDEGKDPAVIRSLTLEPDIIVPGNVLTSVWGTSVPLSPKVDLVEKEVAGLWKIPCTDYIGSCTFHFCDVLDMLIPTGCEPCEPLRTYGLPCHCFKPGEGTSLPKSEFVFPDLELP

SWLTGNYRIEVSLSGSKRLGICIKIAASLKGI"

91..159

sig_peptide

mat_peptide

160..669

/product="G-M2 activator protein"

```

CQ728078      CQ728078      1045 bp      DNA      linear      PAT 03-FEB-2004
LOCUS          Sequence 14012 from Patent WO02068579.
ACCESSION      CQ728078
VERSION        CQ728078.1  GI:42295943
KEYWORDS
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
REFERENCE      Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
AUTHORS        Kits, such as nucleic acid arrays, comprising a majority of
TITLE          humanexons or transcripts, for detecting expression and other uses
JOURNAL        Patent: WO 02068579-A 14012 06-SEP-2002;
FEATURES       PE Corporation (NY) (US)
source         Location/Qualifiers
1. .1045
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:      2.86e-05      Length:      1045
Score:          79.00      Matches:      15
Percent Similarity: 93.75%      Conservative: 0
Best Local Similarity: 93.75%      Mismatches: 1
Query Match:    95.18%      Indels:      0
DB:             6      Gaps:        0

US-10-030-937-72 (1-16) x CQ728078 (1-1045)
Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 524 TACTCACTGCCAAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 571

RESULT 5
HSGM2APB
LOCUS          H.sapiens RNA for GM2-activator protein (clones pGAP2 & pGAP3).
ACCESSION      X61095
VERSION        X61095.1  GI:31856
KEYWORDS       G(M2) activator protein.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE          1 (bases 1 to 1093)
JOURNAL        Nagatani, S., Chen, H.C., Li, S.C., Li, Y.T. and Lockyer, J.M.
MEDLINE        Evidence for two cDNA clones encoding human GM2-activator protein
PUBMED         Biochem. J. 282 (Pt 3), 807-813 (1992)
155364
REFERENCE      2 (bases 1 to 1093)
AUTHORS        Lockyer, J.
TITLE          Direct Submission
JOURNAL        Submitted (26-JUL-1991) J. Lockyer, Tulane University Medical
70112, USA
school, Human Genetics Program, 1430 Tulane Ave., New Orleans LA
FEATURES       Location/Qualifiers
1. .1093
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="pGAP2 and pGAP3"
/tissue_type="placenta"
/clone_lib="placenta lambda gt11"
<1. .603
/function="activator of GM2 hydrolysis"
/codon_start=1
/product="GM2-activator protein"

CDS

```

```

/protein_id="CAA43408.1"
/db_xref="GI:31857"
/db_xref="TReMBL:Q14428"
/translation="RAGPPFPMQSLMOAPLLIALGLLAAQAHLKPKPSQSSFSWD
NCDEGKDPAVIRSLTLRPDPIIVGNVTLVWGSTSVLSPLKVDLVLEKAVGLWI
KIPCTDYIGSCTFEHFCVDLMDLIPTGEPCEPPLRTYGLPCHCPFKEGTSLPKSEFV
VPDLFLPSWLTITGNRYIESVLSSSGKRLGICIKIAASUKGI"
121. .600
/product="unnamed"

mat_peptide
ORIGIN
Alignment Scores:
Pred. No.:      3e-05      Length:      1093
Score:          79.00      Matches:      15
Percent Similarity: 93.75%      Conservative: 0
Best Local Similarity: 93.75%      Mismatches: 1
Query Match:    95.18%      Indels:      0
DB:             9      Gaps:        0

US-10-030-937-72 (1-16) x HSGM2APB (1-1093)
Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 454 TACTCACTGCCAAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 501

RESULT 6
HSGM2A3
LOCUS          Homo sapiens GM2 activator protein (GM2A) gene, exons 3 and 4,
DEFINITION      complete sequence, and complete cds.
ACCESSION      AF124719
VERSION        AF124719.1  GI:4587477
KEYWORDS
SEGMENT
SOURCE         3 of 3
ORGANISM       Homo sapiens (human)
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE          1 (bases 1 to 1706)
JOURNAL        Chen, B., Rigat, B., Curry, C. and Mahuran, D.J.
MEDLINE        Structure of the GM2A gene: identification of an exon 2 nonsense
PUBMED         mutation and a naturally occurring transcript with an in-frame
10364519        deletion of exon 2
REFERENCE      Am. J. Hum. Genet. 65 (1), 77-87 (1999)
2 (bases 1 to 1706)
AUTHORS        Chen, B., Rigat, B., Curry, C. and Mahuran, D.J.
TITLE          Direct Submission
JOURNAL        Submitted (29-JAN-1999) Structural Biology and Biochemistry,
M5G1X8, Canada  Hospital For Sick Children, 555 University Ave., Toronto, Ontario
FEATURES       Location/Qualifiers
1. .1706
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
order(AFI24717.1:1. .633,AF124718.1:1. .1047,1. .1706)
/gene="GM2A"
join(AFI24717.1:1. .171,AF124718.1:391. .552,616. .798,
1180. .>1335)
/gene="GM2A"
/product="GM2 activator protein"
join(AFI24717.1:91. .171,AF124718.1:391. .552,616. .798,
1180. .1335)
/gene="GM2A"
/codon_start=1
/product="GM2 activator protein"
/protein_id="AAD25741.1"
/db_xref="GI:4587479"
/translaton="MQSLMOAPLLIALGLLTPAQAHLKPKPSQSSFSFWDNCDEGKD
PAVIRSLTLRPDPIIVGNVTLVWGSTSVLSPLKVDLVLEKAVGLWIKIPCTDY
IGSCIFEHFCDVLDMLIPTGEPCEPPLRTYGLPCHCPFKEGTSLPKSEFVVPDLFL

```

intron
exon
intron
exon

SWLTTGNYRIESVSSGKRLGCIKIAASLKGI"
1. .615
/gene="GM2A"
/number=2
616. .798
/gene="GM2A"
/number=3
799. .1179
/gene="GM2A"
/number=3
1180. .21335
/gene="GM2A"
/number=4

Alignment Scores:
Pred. No.: 4,77e-05 Length: 1706
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-72 (1-16) x HSGM2A3 (1-1706)

QY 1 TyrSerLeuProlySerGluPheAlaValProAspLeuGluLeuPro 16
LOCUS BC009273
DEFINITION Homo sapiens GM2 ganglioside activator, mRNA (CDNA clone MGC:10462
IMAGE:4053681), complete cds.
DB 1186 TACTCACTGCCAAGAGCGAATTGTTGTCCTGACCTGGAGCTGCC 1233

RESULT 7
BC009273 2413 bp mRNA linear PRI 29-JUN-2004
LOCUS BC009273
DEFINITION Homo sapiens GM2 ganglioside activator, mRNA (CDNA clone MGC:10462
IMAGE:4053681), complete cds.
DB 1186 TACTCACTGCCAAGAGCGAATTGTTGTCCTGACCTGGAGCTGCC 1233

ACCESSION BC009273.2 GI:38197023
VERSION BC009273
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2413)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Udin,T.B., Toshiyuki,S.,
Carrincci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McSwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettaman,M., Madan,A., Rodighiero,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.O., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)
12477932
2 (bases 1 to 2413)
Strausberg,R.
Direct Submission
Submitted (12-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Nov 6, 2003 this sequence version replaced gi:14424506.

REMARK
COMMENT

intron
exon
intron
exon

Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: misc_mgc@ngri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
Series: IRAL Plate: 15 Row: j Column: 22
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 16507969.

FEATURES
Location/Qualifiers
1. .2413
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:10462 IMAGE:4053681"
/tissue_type="Uterus, leiomyosarcoma"
/clone_lib="NIH MGC 46"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. .2413
/gene="GM2A"
/note="synonym: SAP-3"
/db_xref="LocusID:2760"
/db_xref="MIM:272750"
30. .611
/gene="GM2A"
/codon_start=1
/product="GM2 ganglioside activator, precursor"
/protein_id="AAH09273.1"
/db_xref="GI:14424507"
/db_xref="LocusID:2760"
/db_xref="MIM:272750"
/translation="MGSLLQAPLLIALGLLLAQAHLKPSQSLSSFWDCDEKDY
PAVIRSLTLEPDIIVVGNVTLSVVGSTVPLSSPLKVDLVLEKEVAGLWIKPTCTDY
IGSCFPEHFDVLDMLIPTGEPCEPLRTYGLPCHCPKEGTYSLPKSFVFPVDPLELP
SWLTTGNYRIESVSSGKRLGCIKIAASLKGI"

ORIGIN
Alignment Scores:
Pred. No.: 6.85e-05 Length: 2413
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 9 Gaps: 0
US-10-030-937-72 (1-16) x BC009273 (1-2413)
QY 1 TyrSerLeuProlySerGluPheAlaValProAspLeuGluLeuPro 16
DB 462 TACTCACTGCCAAGAGCGAATTGTTGTCCTGACCTGGAGCTGCC 509

RESULT 8
AX330938
LOCUS AX330938
DEFINITION Sequence 1447 from Patent WO0194629.
ACCESSION AX330938
VERSION AX330938.1 GI:18121572

AX330938 2436 bp DNA linear PAT 09-JAN-2002

```

KEYWORDS      Homo sapiens (human)
SOURCE        Homo sapiens
ORGANISM      Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE     Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
AUTHORS       Horrigan, S., Soppet, D.R., and Weaver, Z.
TITLE        Cancer gene determination and therapeutic screening using signature
JOURNAL       gene sets
JOURNAL       Patent: WO 0194629-A 1447 13-DEC-2001;
FEATURES      Avalon Pharmaceuticals (US)
SOURCE        Location/Qualifiers
1. .2436
   /organism="Homo sapiens"
   /mol_type="unassigned DNA"
   /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:      6.92e-05      Length:      2436
Score:          79.00         Matches:     15
Percent Similarity: 93.75%      Conservative: 0
Best Local Similarity: 93.75%    Mismatches:  1
Query Match:    95.18%         Indels:      0
DB:             Gaps:         0

US-10-030-937-72 (1-16) x AX330938 (1-2436)
QY      1 TytSerLeuProlySerGluPheAlaValProAspLeuGluLeuPro 16
Db      491 TACTCACTGCCAAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 538

RESULT 9
HSGM2APT
LOCUS      HSGM2APT      2436 bp      mRNA      linear      PRI 15-FEB-1995
DEFINITION H. sapiens mRNA for GM2 activator protein.
ACCESSION  X62078
VERSION     X62078.1 GI:313158
KEYWORDS    G(M2) activator protein.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 2436)
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
            Klima, H., Tanaka, A., Schnabel, D., Nakano, T., Schroder, M., Suzuki, K.
            and Sandhoff, K.
TITLE       Characterization of full-length cDNAs and the gene coding for the
JOURNAL     human GM2 activator protein
MEDLINE     FEBS Lett. 289 (2), 260-264 (1991)
PUBMED      92008637
REFERENCE   2 (bases 1 to 2436)
AUTHORS     Sandhoff, K.
TITLE       Over-expression of a functionally active human GM2-activator
JOURNAL     protein in Escherichia coli
MEDLINE     Biochem. J. 292 (Pt 2), 571-576 (1993)
PUBMED      93277527
FEATURES    9503891
SOURCE      Location/Qualifiers
1. .2436
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /isolate="patient with juvenile form of Sandhoff disease"
   /db_xref="taxon:9606"
   /clone="pUC18"
   /cell_type="fibroblast"
   /clone_lib="cDNA"
   59. .640
   /note="alternative"
   /codon_start=1
   /product="GM2 activator protein"

CDS

```

```

/protein_id="CAA43993.1"
/db_xref="GI:673415"
/db_xref="GOA:PI7900"
/db_xref="Swiss-Prot:PI7900"
/translation="MQSLMOAPLLIALGLLALTPAQAHKKPSQSLSSFSWDCDEGKD
PAVIRSLTLEPDPPIVVPNGVTLVSGVSTPLSPKLVDLVLEKEVAGLWIKIPCTDY
IGSCTREHPCDVLDMILPTGEPCEPELRITVGLCHCFKPGTYSLPKSEFVVPDLPLP
SWITIGNYRIESVLSSSGKRLGCIKIAASLKI"
71. .640
/note="alternative"
/codon_start=1
/product="GM2 activator protein"
/protein_id="CAA43994.1"
/db_xref="GI:673416"
/translation="MQAPLLIALGLLALTPAQAHKKPSQSLSSFSWDCDEGKDPYVI
RSLTLEPDPPIVVPNGVTLVSGVSTPLSPKLVDLVLEKEVAGLWIKIPCTDYIGSC
TFEHFCDVLDMILPTGEPCEPELRITVGLCHCFKPGTYSLPKSEFVVPDLPLPSWIT
TGNVRIESVLSSSGKRLGCIKIAASLKI"

ORIGIN
Alignment Scores:
Pred. No.:      6.92e-05      Length:      2436
Score:          79.00         Matches:     15
Percent Similarity: 93.75%      Conservative: 0
Best Local Similarity: 93.75%    Mismatches:  1
Query Match:    95.18%         Indels:      0
DB:             Gaps:         0

US-10-030-937-72 (1-16) x HSGM2APT (1-2436)
QY      1 TytSerLeuProlySerGluPheAlaValProAspLeuGluLeuPro 16
Db      491 TACTCACTGCCAAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 538

RESULT 10
AC011391
LOCUS      AC011391      107320 bp      DNA      linear      HTG 23-APR-2001
DEFINITION Homo sapiens chromosome 5 clone CTB-176L22, WORKING DRAFT SEQUENCE,
10 ordered pieces.
ACCESSION  AC011391
VERSION     AC011391.5 GI:9256281
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 107320)
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
            DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 5
JOURNAL     Unpublished
PUBMED      2 (bases 1 to 107320)
REFERENCE   DOE Joint Genome Institute.
AUTHORS     Direct Submission
TITLE       Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
JOURNAL     Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:7710539.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
Project Information
Center Project Name: 132074, H341
Center clone name: CIT978SKB_176L22
-----
Summary Statistics
Consensus quality: 98512 bases at least Q40
Consensus quality: 104463 bases at least Q30
Consensus quality: 105879 bases at least Q20
Estimated insert size: 120000; pulse field gel estimation
Estimated insert size: 106870; sum-of-contigs estimation
Quality coverage: 6.0 in Q20 bases; pulse field gel estimation
Quality coverage: 6.73 in Q20 bases; sum-of-contigs estimation.

```

* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1 7321: contig of 7321 bp in length

7322 7421: gap of unknown length

7422 16470: contig of 9049 bp in length

16471 16570: gap of unknown length

16571 38101: contig of 21531 bp in length

38102 38201: gap of unknown length

38202 54501: contig of 16300 bp in length

54502 54601: gap of unknown length

54602 58262: contig of 3661 bp in length

58263 58362: gap of unknown length

58363 62481: contig of 4119 bp in length

62482 62581: gap of unknown length

62582 77487: contig of 14906 bp in length

77488 77587: gap of unknown length

77588 83728: contig of 6141 bp in length

83729 83828: gap of unknown length

83829 98662: contig of 14834 bp in length

98663 98762: gap of unknown length

98763 107320: contig of 8558 bp in length.

Location/Qualifiers

1. .107320

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="5"

/clone="CTB-176L22"

/clone_lib="CalTech human BAC library B"

ORIGIN

1. .107320

Alignment Scores:

Pred. No.: 0.00359 Length: 107320

Score: 79.00 Matches: 15

Percent Similarity: 93.75% Conservative: 0

Best Local Similarity: 93.75% Mismatches: 1

Query Match: 95.18% Indels: 0

DB: 2 Gaps: 0

US-10-030-937-72 (1-16) x AC011391 (1-107320)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16

Db 29983 TACTCAGTCCCAAGAGCAATTCGTTGTGCTGACCTGGAGCTGCC 30030

RESULT 11

AC011342/c

LOCUS 120584 bp DNA linear PRI 24-OCT-2001

DEFINITION Homo sapiens chromosome 5 clone CTC-276H5, complete sequence.

AC011342

AC011342.5 GI:16356867

VERSION HTG.

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 120584)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE

2 (bases 1 to 120584)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

3 (bases 1 to 120584)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (24-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Oct 24, 2001 this sequence version replaced gi:9256276. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center www.shgc.stanford.edu

Quality: Phrap Quality >=40 99.6% of Sequence; Estimated Total Number of Errors is 0.3.

FEATURES

Location/Qualifiers

1. .120584

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="5"

/clone="CTC-276H5"

ORIGIN

1. .120584

Alignment Scores:

Pred. No.: 0.00406 Length: 120584

Score: 79.00 Matches: 15

Percent Similarity: 93.75% Conservative: 0

Best Local Similarity: 93.75% Mismatches: 1

Query Match: 95.18% Indels: 0

DB: 9 Gaps: 0

US-10-030-937-72 (1-16) x AC011342 (1-120584)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16

Db 11370 TACTCAGTCCCAAGAGCAATTCGTTGTGCTGACCTGGAGCTGCC 11323

RESULT 12

AC008385/c

LOCUS 151712 bp DNA linear PRI 04-JUN-2002

DEFINITION Homo sapiens chromosome 5 clone CTC-224D3, complete sequence.

AC008385

AC008385.7 GI:21321772

VERSION HTG.

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 151712)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE

2 (bases 1 to 151712)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

3 (bases 1 to 151712)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (26-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

4 (bases 1 to 151712)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (04-JUN-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Jun 4, 2002 this sequence version replaced gi:14550298. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center www.shgc.stanford.edu

Quality: Phrap Quality >=40 99.7% of Sequence; Estimated Total Number of Errors is 0.3.

FEATURES

Location/Qualifiers

1. .151712

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-224D3"

ORIGIN
Alignment Scores:
Pred. No.: 0.00515 Length: 151712
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-72 (1-16) x AC008385 (1-151712)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 92074 TACTCACTGCCAAGAGCGAATTCGTTGCTGACCTGAGCTGCC 92027

RESULT 13
LOCUS AB083313 950 bp mRNA linear PRI 29-OCT-2003
DEFINITION Macaca fascicularis gm2a mRNA for ganglioside GM2 activator, complete cds.
ACCESSION AB083313
VERSION AB083313.1 GI:23574732
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Macaca fascicularis (crab-eating macaque)
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.
REFERENCE 1 Kusuda,J., Osada,N., Hida,M., Sugano,S. and Hashimoto,K. Isolation and characterization of cDNA for macaque neurological disease genes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 950) Kusuda,J. Direct Submission
AUTHORS Submitted (08-APR-2002) Jun Kusuda, National Institute of Infectious Diseases, Division of Genetic Resources; Toyama, Shinjuku, Tokyo 1628640, Japan [E-mail:jkusuda@nih.go.jp, URL:http://www.nih.go.jp, Tel:81-3-5285-1111(ex.2122), Fax:81-3-5285-1181]
FEATURES
source
1..950 Location/Qualifiers
/organism="Macaca fascicularis"
/mol_type="mRNA"
/db_xref="taxon:9541"
/clone="QcCE-17591"
/tissue_type="brain cerebellum cortex"
/clone_lib="macaque brain library QcCE"
/note="vector:TOP10"
gene 1..950
/gene="gm2a"
CDS 66..638
/codon_start=1
/product="ganglioside GM2 activator"
/db_xref="GI:23574733"
/db_xref="BAC20592.1"
/translation="MQSLMQAPVLIAGLILFAAPAQAHLLKLGFSWDCDEGKDPV
IRSLTEPPDILIPGNVTVSGVTSVPLSSPLKVLVLEKEVAGWIKIPCTDYIGS
CTFEDSCDVLMDLIPTEGCEPELRTYGLPCHCPFKEGTYSLPKSEFVPHLELPWL
TTGNVYIESILSNRKLKLGICIKIAASLKG"

ORIGIN
Alignment Scores:
Pred. No.: 0.000586 Length: 950
Score: 72.00 Matches: 14

```

```

Percent Similarity: 87.50% Conservative: 0
Best Local Similarity: 87.50% Mismatches: 2
Query Match: 86.75% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-72 (1-16) x AB083313 (1-950)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 489 TACTCACTGCCAAGAGCGAATTCGTTGCTGACCTGAGCTGCC 536

RESULT 14
LOCUS AB051391 600 bp mRNA linear ROD 15-NOV-2002
DEFINITION Rattus norvegicus rGM2AP mRNA for GM2 activator protein, complete cds.
ACCESSION AB051391
VERSION AB051391.1 GI:25006236
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 Miwa,N., Okada,T. and Nakamura,S. GM2AP in hematopoietic cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 600) Miwa,N., Okada,T. and Nakamura,S. Direct Submission
AUTHORS Submitted (20-NOV-2000) Noriko Miwa, Kobe University, 2nd Biochemistry, School of Medicine; Chuo-ku, Kusunoki-cho, 7-5-1, Kobe-city, Hyogo 650-0017, Japan [E-mail:miwanori@post.med.kobe-u.ac.jp, Tel:81-78-382-5421, Fax:81-78-382-5439]
FEATURES
source
1..600 Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
gene 1..600
/gene="rGM2AP"
CDS 1..600
/gene="rGM2AP"
/codon_start=1
/product="GM2 activator protein"
/protein_id="BAC24018.1"
/db_xref="GI:25006237"
/translation="MRRVPELLVLGLLFLVGLFAGVPAPSLIKRPSQLQGFSDN
CDEGKDPVILKLTLPDPFIVPGDVIVSAGKTSIPTSPQKVELTVKEVAGFWVK
IPCVEQLGSCCTYENVCDLIDQYIPGCTCPEPLHTYGLPCHCPFKEGTYSLPSSNPTV
PDLELPSSLTGNRYIQSILSSGGKRLKICIKIAASLKG"

ORIGIN
Alignment Scores:
Pred. No.: 0.00138 Length: 600
Score: 69.00 Matches: 13
Percent Similarity: 81.25% Conservative: 0
Best Local Similarity: 81.25% Mismatches: 3
Query Match: 83.13% Indels: 0
DB: 10 Gaps: 0

US-10-030-937-72 (1-16) x AB051391 (1-600)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 451 TACTCACTGCCCTCGAGCAACTTCACAGTGCCTGAGCTGCC 498

RESULT 15
LOCUS EL2286 1983 bp DNA linear PAT 27-APR-1998
DEFINITION cDNA encoding rat GM2 activator protein.

```

ACCESSION E12286
VERSION E12286.1 GI:3251120
KEYWORDS JP 1996308582-A/1.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1983)
AUTHORS Takizawa, M. and Matsuo, N.
TITLE RAT GM2 ACTIVATOR PROTEIN GENE
JOURNAL Patent: JP 1996308582-A 1 26-NOV-1996;
KAO CORP
COMMENT OS Rattus sp.
PN JP 1996308582-A/1
PD 26-NOV-1996
PF 23-MAY-1995 JP 1995123757
PI TAKIZAWA MINORU, MATSUO NOBORU
PC C12N15/09, C07H21/04//C07K14/47, C12N1/21, C12P21/02, (C12N1/21,
C12R1:19),
PC (C12P21/02, C12R1:19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FH
FT source 1..1983
FT /organism='Rattus sp.'
FT /strain='Wistar'
FT /tissue_type='brain'
FT 5'UTR 1..11
FT CDS 12..611
FT /product='rat GM2 activator protein' FT
FT 3'UTR 612..1983.
FEATURES
source
1..1983
Location/Qualifiers
/organism='Rattus sp.'
/mol_type='genomic DNA'
/db_xref='taxon:10118'
ORIGIN
Alignment Scores:
Pred. No.: 0.00481 Length: 1983
Score: 69.00 Matches: 13
Percent Similarity: 81.25% Conservative: 0
Best Local Similarity: 81.25% Mismatches: 3
Query Match: 83.13% Indels: 0
DB: 6 Gaps: 0
US-10-030-937-72 (1-16) x E12286 (1-1983)
QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 462 TACTCACTGCCTTCGAGCAACTTCACAGTGCCTGATCIGGAGCTTCCA 509
Search completed: November 18, 2004, 22:22:17
Job time : 674.378 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 16:05:15 ; Search time 71.9644 Seconds
(without alignments)
1167.114 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 83

Sequence: 1 YSLPKSEFAVPDLELP 16

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US1003937/runat 16112004 153013 2903/app.query.fasta_1.789
-DB=N Geneseq 23Sep04 -Qfmt=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptp -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1003937 -CGEN 1_1_1240 @runat 16112004 153013 2903 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_23Sep04.*

1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	95.2	448	6	Abq60520 Human col
2	79	95.2	953	12	Adq17711 Human sof
3	79	95.2	1043	4	Aaf54705 Nucleotid
4	79	95.2	1043	4	Aaf54708 Nucleotid
5	79	95.2	1705	4	Aaf54707 Nucleotid
6	79	95.2	1706	4	Aaf54701 Nucleotid

7	79	95.2	1706	4	Aaf54704
8	79	95.2	1935	10	ADB47402
9	79	95.2	2384	6	ABK34915
10	79	95.2	2436	3	AAC55714
11	79	95.2	2436	6	ABL63110
12	79	95.2	2436	10	ADD71046
13	79	95.2	2436	11	ADN95859
14	79	95.2	2471	5	AAS64907
15	79	95.2	2478	6	ABV78068
16	79	95.2	2478	12	ADN03619
17	79	95.2	2498	5	AAS81113
18	79	95.2	3988	12	ADO22367
19	79	95.2	249999	8	ABZ80229
20	69	83.1	1983	2	AAT61025
21	69	83.1	1983	10	ADB52361
22	60	72.3	48	4	Aaf54726
23	60	72.3	579	4	Aaf54730
24	56	67.5	579	4	Aaf54698
25	53	63.9	1098	5	AAS84304
26	47	56.6	588	8	ACF75089
27	47	56.6	790	3	AAF12029
28	47	56.6	4210	4	ABL10029
29	47	56.6	4318	4	ABL21121
30	47	56.6	8220	2	AAQ83529
31	47	56.6	8220	2	AAQ83529
32	47	56.6	8220	3	AAZ98286
33	47	56.6	9280	2	AAV74442
34	47	56.6	10549	5	ABAI5442
35	47	56.6	13713	4	ABL21120
36	47	56.6	13749	4	ABL10028
37	47	56.6	19124	2	AAQ72882
38	47	56.6	19124	3	AAZ98287
39	47	56.6	29871	6	ABN86359
40	47	56.6	68230	9	ADA66349
41	47	56.6	68233	9	ADA03065
42	47	56.6	68233	10	ADB72803
43	47	56.6	68233	11	ADL27143
44	47	56.6	110000	9	ACH03408
45	46	55.4	1195	5	AAS75072

ALIGNMENTS

RESULT 1
ABQ60520
ID ABQ60520 standard; cDNA; 448 BP.
XX
AC ABQ60520;
XX
DT 02-AUG-2002 (first entry)
XX
DE Human colon cancer related nucleotide sequence SEQ ID NO:4215.
XX
KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KW genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200229086-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US030732.
XX
PR 02-OCT-2000; 2000US-0237271P.
XX
PA (FARB) BAYER CORP.
XX
PI Burgess C, Asle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
XX Thiaglingam A, Lewis ME;
XX WPI; 2002-426115/45.
XX

PT New isolated nucleic acid that is differentially expressed in cancer
PT tissues useful for determining the presence of colon cancer in a cell or
PT tissue type, and in antisense therapy.

XX Claim 1; Fig 1; 796pp; English.

XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC used in antisense therapy. An antibody immunoreactive with a polypeptide
CC encoded by (I) is useful for detecting cancer in a patient sample, and
CC for detecting the presence or absence of a polynucleotide encoded by a
CC nucleic acid which hybridizes to (I) in a cell. A probe/primer derived
CC from (I) can be used for determining the presence of a nucleic acid which
CC hybridizes to (I), and for determining the phenotype of cells in a sample
CC of cells from a patient. (I) is useful for determining the presence of
CC colon cancer in a cell or tissue type, for determining the presence or
CC state of other type of cancer, in antisense therapy, to generate
CC macroarrays on a solid surface, to identify a chromosome on which the
CC corresponding gene resides, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (I) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists
XX
SQ Sequence 448 BP; 98 A; 118 C; 101 G; 110 T; 0 U; 21 Other;

Alignment Scores:
Pred. No.: 6.26e-05 Length: 448
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-72 (1-16) x ABQ60520 (1-448)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
DB 43 TACTCAGTCCCAAGAGCGAATTCGTTGGCTGACCTGGAGTGCCTGCC 90

RESULT 2

ADQ17711
ID ADQ17711 standard; DNA; 953 BP.
XX
AC ADQ17711;
XX
XX 26-AUG-2004 (first entry)
XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 528.
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX Homo sapiens.
XX
XX WO2004048938-A2.
XX
XX 10-JUN-2004.
XX
XX 26-NOV-2003; 2003WO-US038193.
XX
XX 26-NOV-2002; 2002US-0429739P.
XX
XX (PROT-) PROTEIN DESIGN LABS INC.
XX
XX Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX

PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX

PS Example 2; SEQ ID NO 528; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

SQ Sequence 953 BP; 201 A; 286 C; 231 G; 235 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.000156 Length: 953
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 12 Gaps: 0

US-10-030-937-72 (1-16) x ADQ17711 (1-953)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
DB 523 TACTCAGTCCCAAGAGCGAATTCGTTGGCTGACCTGGAGTGCCTGCC 570

RESULT 3

AAF54705
ID AAF54705 standard; DNA; 1043 BP.
XX
AC AAF54705;
XX
XX 15-MAY-2001 (first entry)
XX
XX Nucleotide sequence of a human polynucleotide sequence.
XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.
XX Homo sapiens.
XX
XX WO200105422-A2.
XX
XX 25-JAN-2001.
XX
XX 17-JUL-2000; 2000WO-FR002057.
XX
XX 15-JUL-1999; 99FR-00009372.
XX
XX (INMR) BIOMERIEUX STELHYS.
XX
XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX

PS Claim 11; Page 181-182; 209pp; French.
XX
XX The present sequence represents a human polynucleotide sequence, which is
XX used in the method of the invention. The specification describes a method
XX which uses at least one polypeptide or polynucleotide sequence belonging

PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX

CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX

SQ Sequence 1043 BP; 222 A; 303 C; 247 G; 271 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.000174 Length: 1043
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-72 (1-16) x AAF54705 (1-1043)

QY 1 TyrSerLeuProLysSerGlupheAlaValProAspLeuGluLeuPro 16
 DB 523 TACTCAGTCCCAAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 570

RESULT 4
 AAF54708
 ID AAF54708 standard; DNA; 1043 BP.
 AC AAF54708;
 XX

XX 15-MAY-2001 (first entry)
 XX Nucleotide sequence of a human polynucleotide sequence.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.
 XX Homo sapiens.

XX WO200105422-A2.
 XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.
 XX 15-JUL-1999; 99FR-00009372.

XX (INMR) BIOMERIEUX STELHYS.
 XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.
 XX Detecting, preventing and treating degenerative, neurological and

PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX Claim 11; Page 183; 209pp; French.

XX The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or

CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX

SQ Sequence 1043 BP; 222 A; 303 C; 247 G; 271 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.000174 Length: 1043
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-72 (1-16) x AAF54708 (1-1043)

QY 1 TyrSerLeuProLysSerGlupheAlaValProAspLeuGluLeuPro 16
 DB 523 TACTCAGTCCCAAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 570

RESULT 5
 AAF54707
 ID AAF54707 standard; DNA; 1705 BP.
 AC AAF54707;
 XX

XX 15-MAY-2001 (first entry)
 XX Nucleotide sequence of a human polynucleotide sequence.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.
 XX Homo sapiens.

XX WO200105422-A2.
 XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.
 XX 15-JUL-1999; 99FR-00009372.

XX (INMR) BIOMERIEUX STELHYS.
 XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.
 XX Detecting, preventing and treating degenerative, neurological and

PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX Claim 11; Page 182-183; 209pp; French.

XX The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and

CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX
 SQ Sequence 1705 BP; 427 A; 390 C; 412 G; 476 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.000314 Length: 1705
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservatives: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-72 (1-16) x AAF54707 (1-1705)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
 ID AAF54701 standard; DNA; 1706 BP.
 DB 1185 TACTCACTGCCAAGAGCGAATTGTTGTGCTGACCTGGAGCTGCC 1232

RESULT 6
 AAF54701
 ID AAF54701 standard; DNA; 1706 BP.

XX AC AAF54701;

XX DT 15-MAY-2001 (first entry)

XX DE Nucleotide sequence of a human polynucleotide sequence.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99FR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX DR WPI; 2001-159475/16.

XX PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX PS Claim 11; Page 179-180; 209pp; French.

XX The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in

CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX
 SQ Sequence 1706 BP; 427 A; 391 C; 412 G; 476 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.000314 Length: 1706
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservatives: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-72 (1-16) x AAF54701 (1-1706)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
 DB 1186 TACTCACTGCCAAGAGCGAATTGTTGTGCTGACCTGGAGCTGCC 1233

RESULT 7
 AAF54704
 ID AAF54704 standard; DNA; 1706 BP.

XX AC AAF54704;

XX DT 15-MAY-2001 (first entry)

XX DE Nucleotide sequence of a human polynucleotide sequence.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99FR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX DR WPI; 2001-159475/16.

XX PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX PS Claim 11; Page 181; 209pp; French.

XX The present sequence represents a human polynucleotide sequence, which is
 CC used in the method of the invention. The specification describes a method
 CC which uses at least one polypeptide or polynucleotide sequence belonging
 CC to the perlecan, precursor of the retinol-binding plasma protein,
 CC precursor of the ganglioside GM2 activator, calgranulin B or saposin B
 CC protein families. The method is used for detecting, preventing or
 CC treating a degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX SQ Sequence 1706 BP; 427 A; 391 C; 412 G; 476 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.000314 Length: 1706
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-72 (1-16) x AAF54704 (1-1706)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
 Db 1186 TACTCACTGCCAAGAGCGAATTCGTTGCTGACCTGGAGCTGCC 1233

RESULT 8
 ADB47402
 ID ADB47402 standard; cDNA; 1935 BP.
 XX AC ADB47402;
 XX DT 04-DEC-2003 (first entry)
 XX DE Human cDNA upregulated in dendritic cells SEQ ID NO 102.
 XX KW ss: gene; human; dendritic cells; high throughput; cancer;
 KW infectious disease; autoimmune disease; allergy;
 KW graft versus host disease; vaccine enhancing; gene therapy.
 XX OS Homo sapiens.
 XX PN US2003134283-A1.
 XX PD 17-JUL-2003.
 XX PF 03-OCT-2001; 2001US-00971392.
 XX PR 03-OCT-2000; 2000US-0237652P.
 XX PA (PETE/) PETERSON D P.
 PA (PEAR/) PEARSON C I.
 PA (COCK/) COCKS B G.
 XX PI Peterson DP, Pearson CI, Cocks BG;
 XX WPI; 2003-662509/62.
 XX New combination comprises cDNAs that are differentially expressed in dendritic cells useful for preparing a composition for diagnosing or treating cancer, infectious disease, autoimmune, allergy or graft versus host disease.
 XX Claim 1; SEQ ID NO 102; 28pp; English.
 XX The invention relates to a combination comprising cDNAs that are differentially expressed in dendritic cells (DC). Also included is a high throughput method for detecting differential expression of one or more cDNAs in a sample containing nucleic acids. The combination is useful for preparing a composition for diagnosing, treating and monitoring the treatment of cancer, infectious disease, autoimmune, allergy or graft versus host disease, or for enhancing a vaccine. The present sequence represents a human cDNA upregulated in dendritic cells. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030134283.
 XX SQ Sequence 1935 BP; 459 A; 507 C; 460 G; 509 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.000366 Length: 1935
 Score: 79.00 Matches: 15

Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 10 Gaps: 0

US-10-030-937-72 (1-16) x ADB47402 (1-1935)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
 Db 534 TACTCACTGCCAAGAGCGAATTCGTTGCTGACCTGGAGCTGCC 581

RESULT 9
 ABK34915
 ID ABK34915 standard; cDNA; 2384 BP.
 XX AC ABK34915;
 XX DT 08-MAY-2002 (first entry)
 XX DE Human cDNA encoding secreted protein #53.
 XX KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
 KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
 KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
 KW tissue regeneration; wound healing; burn; haematopoiesis;
 KW myeloid cell deficiency; lymphoid cell deficiency.
 XX OS Homo sapiens.
 XX PN WO200177288-A2.
 XX PD 18-OCT-2001.
 XX PF 29-MAR-2001; 2001WO-US010224.
 XX PR 06-APR-2000; 2000US-0195582P.
 XX PA (GEMY) GENETICS INST INC.
 XX PI Wong GG, Clark HF, Pechtel K, Agostino MJ, Howes SH, Resnick RJ;
 PI Gulukota K, Graham JR;
 XX WPI; 2002-179321/23.
 XX Five hundred and ninety two polynucleotides derived from a variety of human tissue sources which encode secreted proteins, useful for treating immune deficiencies and disorders such as autoimmune disorders.
 XX Claim 1; Page 95-96; 372pp; English.
 XX The invention relates to 592 polynucleotides which have been derived from a variety of human tissue sources and which encode novel secreted proteins. The polynucleotides can be used as probes for the identification and isolation of full length cDNA and genomic DNA. The polynucleotides and proteins can also be used as nutritional supplements. The proteins are useful in the treatment of various immune deficiencies and disorders such as viral infections, bacterial infections, fungal infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions and conditions (e.g. asthma). They are also useful for treating neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's disease), liver fibrosis, coagulation disorders (e.g. haemophilia), inflammatory disorders (e.g. Crohn's disease) and tumours. They are also useful for tissue regeneration, for wound healing and in the treatment of burns, incisions and ulcers. The proteins are also useful for regulating haematopoiesis and for treating myeloid or lymphoid cell deficiencies. Sequences ABK34863-ABK35454 represent polynucleotides of the invention
 XX SQ Sequence 2384 BP; 550 A; 618 C; 571 G; 645 T; 0 U; 0 Other;

Fri Nov 19 14:12:37 2004

us-10-030-937-72.p2n.rng

Alignment Scores:
Pred. No.: 0.000471 Length: 2384
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-72 (1-16) x ABK34915 (1-2384)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 445 TACTCACTGCCCAAGAGCGAATTCTGTGCTGACCTGGAGCTGCC 492

RESULT 10

AAC55714
ID AAC55714 standard; cDNA; 2436 BP.

XX AC AAC55714;

XX DT 17-JAN-2001 (first entry)

XX DE Human GM2 activator protein cDNA sequence from Genbank X62078.

XX KW Human; differentially regulated gene; macrophage development; diagnosis;
KW matrix metalloproteinase 19; MMP19; antiarthritic; antiinflammatory;
KW destructive macrophage development inhibitor; arthritis;
KW colorectal cancer; immune response; ss.

XX OS Homo sapiens.

XX PN WO200055373-A2.

XX PD 21-SEP-2000.

XX PF 15-MAR-2000; 2000WO-US006883.

XX PR 15-MAR-1999; 99US-0124530P.

XX PA (EOSB-) EOS BIOTECHNOLOGY INC.

XX PI Murray R;

XX DR WPI; 2000-628200/60.

XX PT Screening drug candidates comprising adding a drug to a cell expressing an
XX expression profile gene and determining the effect of the drug on the
XX expression of the expression profile gene.

XX PS Claim 1; Page; 99pp; English.

XX CC The present invention describes a method for screening drug candidates.
XX The method comprises adding a drug to a cell that expresses an expression
XX profile gene encoding a protein encoded by 5 sequences of defined base
XX pairs as given in C55638, C55642, C55643, C55644 and C55653 or a sequence
XX represented by Genbank accession number X92521, X62466, J04130, X62087
XX and X76534 (or a fragment) and determining the effect of the drug on the
XX expression of the expression profile gene. An inhibitor of matrix
XX metalloproteinase 19 (MMP-19), preferably an antibody, is useful for
XX treating destructive macrophage disorders (DMD) by inhibiting DM
XX development in a cell of an individual having arthritis. Antibodies to
XX MMP-19 are useful for localising a therapeutic moiety preferably
XX cytotoxic agent or a radioisotope to colorectal cancer tissue. A
XX composition comprising MMP-19 is useful for eliciting an immune response
XX in an individual. C55635 to C55710 represent human differentially
XX regulated genes of the invention. The present sequence represents the
XX human GM2 activator protein cDNA sequence according to the Genbank
XX accession number X62078. N.B. The present sequence is not given in the
XX present specification, but it is specifically claimed by its Genbank
XX accession number

XX SQ Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.000483 Length: 2436
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 3 Gaps: 0

US-10-030-937-72 (1-16) x AAC55714 (1-2436)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 491 TACTCACTGCCCAAGAGCGAATTCTGTGCTGACCTGGAGCTGCC 538

RESULT 11

ABL63110

ID ABL63110 standard; DNA; 2436 BP.

XX AC ABL63110;

XX DT 15-MAY-2002 (first entry)

XX DE Breast cancer related gene sequence SEQ ID NO:1447.

XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cystostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.

XX OS Homo sapiens.

XX PN WO200194629-A2.

XX PD 13-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US010838.

XX PR 05-JUN-2000; 2000US-0209473P.

XX PR 05-JUN-2000; 2000US-0209531P.

XX PR 18-SEP-2000; 2000US-0233133P.

XX PR 18-SEP-2000; 2000US-0233617P.

XX PR 20-SEP-2000; 2000US-0234009P.

XX PR 20-SEP-2000; 2000US-0234034P.

XX PR 20-SEP-2000; 2000US-0234052P.

XX PR 22-SEP-2000; 2000US-0234509P.

XX PR 22-SEP-2000; 2000US-0234567P.

XX PR 25-SEP-2000; 2000US-0234923P.

XX PR 25-SEP-2000; 2000US-0234924P.

XX PR 25-SEP-2000; 2000US-0235077P.

XX PR 25-SEP-2000; 2000US-0235082P.

XX PR 25-SEP-2000; 2000US-0235134P.

XX PR 25-SEP-2000; 2000US-0235280P.

XX PR 26-SEP-2000; 2000US-0235637P.

XX PR 26-SEP-2000; 2000US-0235638P.

XX PR 27-SEP-2000; 2000US-0235711P.

XX PR 27-SEP-2000; 2000US-0235720P.

XX PR 27-SEP-2000; 2000US-0235840P.

XX PR 28-SEP-2000; 2000US-0235863P.

XX PR 28-SEP-2000; 2000US-0236028P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236033P.

XX PR 28-SEP-2000; 2000US-0236034P.

XX PR 28-SEP-2000; 2000US-0236109P.

XX PR 28-SEP-2000; 2000US-0236111P.

XX PR 29-SEP-2000; 2000US-0236842P.

XX PR 29-SEP-2000; 2000US-0236891P.

XX PR 02-OCT-2000; 2000US-0237172P.

XX PR 02-OCT-2000; 2000US-0237173P.

XX PR 02-OCT-2000; 2000US-0237278P.

XX PR 02-OCT-2000; 2000US-0237294P.

XX PR 02-OCT-2000; 2000US-0237295P.

XX PR 02-OCT-2000; 2000US-0237316P.

XX PR 03-OCT-2000; 2000US-0237425P.

PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX (AVAL-) AVALON PHARM.
 PA Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 XX Soppet DR, Weaver Z;
 PI WPI; 2002-188264/24.
 DR
 XX
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 XX
 XX Claim 1; SEQ ID NO 1447; 44pp; English.
 PS
 XX The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms's
 CC tumour
 XX
 SQ Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.000483 Length: 2436
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 6 Gaps: 0
 US-10-030-937-72 (1-16) x ABL63110 (1-2436)
 QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
 Db 491 TACTCACTGCCCAAGAGCGAATTGTTGTCCTGACCTGGAGCTGCC 538
 RESULT 12
 ADD71046
 ID ADD71046 standard; DNA; 2436 BP.
 XX
 AC ADD71046;
 XX
 XX 15-JAN-2004 (first entry)
 XX
 XX Human GM2 ganglioside activated protein gene SEQ ID NO:50.
 DE
 DE liver cancer; chronic hepatitis; cirrhosis; liver disease; hepatotropic;
 KW cytotostatic; gene therapy; human; gene; ds.
 KW
 XX Homo sapiens.
 OS
 XX WO2003061564-A2.
 PN
 XX 31-JUL-2003.
 PD
 XX

PF 20-DEC-2002; 2002WO-US040718.
 XX
 XX 21-DEC-2001; 2001US-0341815P.
 PR 31-DEC-2001; 2001US-0343185P.
 XX
 XX (GENE-) GENE LOGIC INC.
 PA (LGEI-) LG BIOMEDICAL INST.
 PA
 XX Koh SS, Liu Q, Chung H, Zeng W, Lee B, Yeramilli S, Song SY;
 PI WPI; 2003-663343/62.
 DR
 XX
 XX Diagnosing liver cancer cells, useful for treating liver cancer
 PT associated with chronic hepatitis or cirrhosis comprises detecting the
 PT level of expression in a tissue sample of one or more genes associated
 PT with cancerous liver tissues.
 XX
 XX Claim 1; SEQ ID NO 50; 176pp; English.
 PS
 XX The present invention describes a method for diagnosing liver cancer
 CC cells comprising detecting the level of expression in a tissue sample of
 CC one or more genes given in the specification (see ADD70997 to ADD71105),
 CC where differential expression of the genes is indicative of liver cancer.
 CC Also described: (1) detecting the progression of liver cancer in a
 CC patient; (2) monitoring the treatment of a patient with liver cancer; (3)
 CC treating a patient with liver cancer; (4) typing a liver disease in a
 CC patient; (5) detecting the presence or progression of liver cancer in a
 CC patient with chronic hepatitis or cirrhosis; (6) differentiating liver
 CC cancer related to chronic hepatitis from liver cancer related to
 CC cirrhosis; (7) screening for an agent capable of modulating the onset or
 CC progression of liver cancer; (8) a composition comprising at least two
 CC oligonucleotides comprising a sequence that specifically hybridises to
 CC any of the genes; (9) a solid support comprising the at least two
 CC oligonucleotides; (10) a computer system comprising a database containing
 CC information identifying the level in liver tissue of a set of genes; (11)
 CC a method for using the computer system to present information identifying
 CC the expression level in tissue or cell of any of the genes; and (12) a
 CC therapeutic agent for slowing or halting the progression of liver cancer.
 CC The methods are useful for treating liver cancer associated with chronic
 CC hepatitis or cirrhosis. The present sequence represents a specifically
 CC claimed human gene sequence which is used in the exemplification of the
 CC present invention.
 XX
 SQ Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.000483 Length: 2436
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 10 Gaps: 0
 US-10-030-937-72 (1-16) x ADD71046 (1-2436)
 QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
 Db 491 TACTCACTGCCCAAGAGCGAATTGTTGTCCTGACCTGGAGCTGCC 538
 RESULT 13
 ADN95859
 ID ADN95859 standard; DNA; 2436 BP.
 XX
 AC ADN95859;
 XX
 XX 01-JUL-2004 (first entry)
 DT
 XX Human BEC/LEC-related gene sequence SeqID783.
 DE
 XX growth; differentiation; blood endothelial cell; BEC;
 KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
 KW lymphatic growth agent; VEGF-D; antiangiogenic; cytotostatic;
 KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;

KW inflammatory disease; cancer metastasis; lymphatic system; gene; ds;
KW human.
XX Homo sapiens.
XX WO2003080640-A1.
XX 02-OCT-2003.
XX 07-MAR-2003; 2003WO-US006900.
XX 07-MAR-2002; 2002US-0363019P.
XX (LUDW-) LUDWIG INST CANCER RES.
XX (LICN) LICENTIA LTD.
XX Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
XX WPI; 2003-876899/81.
XX P-PSDB; ADN95858.
XX Example 1; SEQ ID NO 783; 176pp; English.
XX This invention relates to a method of differentially modulating the
XX growth or differentiation of blood endothelial cells (BEC) or lymphatic
XX endothelial cells (LEC) comprises contacting endothelial cells with a
XX composition comprising an agent that differentially modulates blood or
XX lymphatic endothelial cells. Treating hereditary lymphoedema comprises
XX identifying a human subject with lymphoedema and with a mutation in at
XX least one allele of a gene encoding a LEC protein, where the mutation
XX correlates with lymphoedema in human subjects, and with the proviso that
XX the LEC protein is not VEGFR-3; and administering to the subject a
XX composition comprising a lymphatic growth agent selected from VEGF-C or
XX VEGF-D polypeptides and polynucleotides. The invention may be useful for
XX the development of compounds with an antiangiogenic, cytostatic,
XX vasotropic or antiinflammatory activity or for gene therapy. The method
XX is useful in modulating the growth or differentiation of blood
XX endothelial cells or lymphatic endothelial cells, in treating hereditary
XX lymphoedema, in screening for an endothelial cell disorder or
XX predisposition to the disorder or in monitoring the efficacy or toxicity
XX of a drug on endothelial cells. The agent is useful in manufacturing a
XX medicament for the differential modulation of blood vessel endothelial
XX cell or lymphatic vessel endothelial cell growth or differentiation. The
XX lymphatic growth agent may also be used in manufacturing a medicament for
XX the treatment of hereditary lymphoedema resulting from a mutation in a
XX LEC gene or of other diseases involving the lymphatic vessels, such as
XX various inflammatory diseases and cancer metastasis via the lymphatic
XX system. The present sequence is that of a human LEC/BEC differentially
XX expressed gene which is related to the method of the invention. Note: This
XX sequence does not appear in the specification but was obtained by the
XX indexer using the source data given in table 14 of the specification.
XX
XX Sequence 2436 BP; 570 A; 632 C; 576 G; 658 T; 0 U; 0 Other;

Alignment Scores: Length: 2436
Pred. No.: 0.000483 Matches: 15
Score: 79.00
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 11 Gaps: 0

US-10-030-937-72 (1-16) x ADN95859 (1-2436)

Qy 1 TvrSxerLeuProlySerGluPheAlaValProAspLeuGluLeuPro 16
Db 491 TACTCAGTGCCTGAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 538

RESULT 14
AAS64907
ID AAS64907 standard; cDNA; 2471 BP.
XX
XX AAS64907;

XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #711.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG00720.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID NO 711; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (II) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activities. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2471 BP; 578 A; 639 C; 583 G; 671 T; 0 U; 0 Other;

Alignment Scores: Length: 2471
Pred. No.: 0.000492 Matches: 15
Score: 79.00
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 5 Gaps: 0

US-10-030-937-72 (1-16) x AAS64907 (1-2471)

Qy 1 TvrSxerLeuProlySerGluPheAlaValProAspLeuGluLeuPro 16
Db 525 TACTCAGTGCCTGAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 572

RESULT 15
ABV78068

ID XX ABV78068 standard; DNA; 2478 BP.
AC XX ABV78068;
XX
DT 12-NOV-2002 (first entry)
XX
DE XX Hypoxia-regulated protein coding sequence #88.
XX
KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;
KW antiinflammatory; vulnery; gynecological; ophthalmological; vaccine;
KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;
KW preeclampsia; atherosclerosis; inflammatory condition; wound healing;
KW inflammation; erythropoiesis; hair loss; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200246465-A2.
XX
PD 13-JUN-2002.
XX
PF 10-DEC-2001; 2001WO-GB005458.
XX
PR 08-DEC-2000; 2000GB-00030076.
PR 08-FEB-2001; 2001GB-00003156.
PR 25-OCT-2001; 2001GB-00025666.
XX
PA (OXFO-) OXFORD BIOMEDICA UK LTD.
XX

White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;
Rayner WN;
WPI; 2002-627238/67.

Identifying a gene involved in disease for treating hypoxia-regulated
conditions, comprises comparing the transcriptome/proteome of two cell
types under different conditions and identifying a differentially
regulated gene.

Claim 37; Page 397-398; 538pp; English.

The present invention relates to methods for identifying genes and
proteins that are implicated in a specific disease or physiological
condition. The method comprises comparing the transcriptome/proteome of a
specialised cell type implicated in a disease or condition with that of a
second specialised cell type, under two experimental conditions, and
identifying a gene that is differentially regulated in the two
specialised cell types under experimental conditions. ABV77873-ABV78116
and ABP65061-ABP65257 were identified using the methods of the invention.
The coding sequences and proteins are useful for treating a disease in a
patient, for manufacture of a medicament for treating hypoxia-regulated
conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
biological response to hypoxia conditions, or hypoxic-associated
pathology in a patient. The coding sequences and proteins are also useful
for monitoring the therapeutic treatment of a disease or physiological
condition, such as cancer, ischaemic conditions, reperfusion injury,
retinopathy, neonatal stress, preeclampsia, atherosclerosis, inflammatory
conditions, wound healing, inflammation, erythropoiesis or hair loss

Sequence 2478 BP; 588 A; 638 C; 586 G; 666 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.000493 Length: 2478
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-72 (1-16) x ABV78068 (1-2478)

QY 1 TyrSerLeuProIysSerGluPheAlaValProAspLeuGluLeuPro 16
|||||

Db 528 TACTCACTGCCCAAGAGCGAATTGCTTGCTGACCTGGAGCTGCC 575

Search completed: November 18, 2004, 19:48:45
Job time : 74.9644 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 19:19:01 ; Search time 13.5822 Seconds
(without alignments)
837.317 Million cell updates/sec

Title: US-10-030-937-72
Perfect score: 83
Sequence: 1 YSLPKSEFAVDLELP 16

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_P/US10030937/runat_16112004_153014_2947/app_query.fasta_1.789
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LCOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937@cgn2_1_200 @runat_16112004_153014_2947 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	56.6	8220	2	US-08-568-459A-11
2	47	56.6	8220	2	US-08-487-826B-11
3	47	56.6	8220	3	US-09-210-288-11
4	47	56.6	9280	4	US-08-956-171E-131
5	47	56.6	9280	4	US-08-781-986A-131
6	47	56.6	19124	4	US-08-487-826B-13
7	46	55.4	1062	4	US-09-894-844-61
8	46	55.4	4403765	3	US-09-103-840A-2
9	46	55.4	4403765	3	US-09-103-840A-2
10	46	55.4	4411529	3	US-09-103-840A-1
11	46	55.4	4411529	3	US-09-103-840A-1
12	44	53.0	924	4	US-09-543-681A-69

13	43	51.8	1416	3	US-08-911-853-3	Sequence 3, Appli
14	43	51.8	1416	3	US-09-479-409-3	Sequence 3, Appli
15	43	51.8	1416	3	US-09-479-453-3	Sequence 3, Appli
16	43	51.8	2439	4	US-09-489-039A-7111	Sequence 7111, Ap
17	43	51.8	4377	3	US-08-911-853-28	Sequence 28, Appli
18	43	51.8	4377	3	US-09-479-409-28	Sequence 28, Appli
19	43	51.8	4377	3	US-09-479-453-28	Sequence 28, Appli
20	43	51.8	112132	4	US-09-741-150-3	Sequence 3, Appli
21	43	51.8	112132	4	US-10-160-187-3	Sequence 3, Appli
22	43	51.8	1830121	4	US-09-557-884-1	Sequence 1, Appli
23	43	51.8	1830121	4	US-09-643-990A-1	Sequence 1, Appli
24	43	51.8	1830121	4	US-10-329-960-1	Sequence 1, Appli
25	42	50.6	354	4	US-09-513-999C-13536	Sequence 13536, A
26	42	50.6	1404	4	US-09-614-221A-277	Sequence 277, App
27	41.5	50.0	2772	3	US-08-936-135-1	Sequence 1, Appli
28	41.5	50.0	2772	4	US-09-439-711C-1	Sequence 1, Appli
29	41.5	50.0	5653	4	US-09-583-638-1	Sequence 1, Appli
30	41	49.4	360	4	US-09-583-110-1456	Sequence 1456, Ap
31	41	49.4	447	4	US-09-252-991A-10231	Sequence 10231, A
32	41	49.4	495	4	US-09-489-039A-3316	Sequence 3316, Ap
33	41	49.4	517	4	US-09-270-767-27920	Sequence 27920, A
34	41	49.4	561	4	US-09-107-532A-1559	Sequence 1559, Ap
35	41	49.4	576	4	US-09-221-017B-842	Sequence 842, App
36	41	49.4	681	4	US-09-107-532A-3105	Sequence 3105, App
37	41	49.4	1017	4	US-09-252-991A-10799	Sequence 10799, A
38	41	49.4	1043	3	US-09-422-576D-6	Sequence 6, Appli
39	41	49.4	1091	3	US-09-422-576D-5	Sequence 5, Appli
40	41	49.4	1101	4	US-09-252-991A-10349	Sequence 10349, A
41	41	49.4	1107	4	US-09-614-221A-398	Sequence 398, App
42	41	49.4	1165	3	US-09-422-576D-25	Sequence 25, Appli
43	41	49.4	1167	4	US-09-270-767-12191	Sequence 12191, A
44	41	49.4	1691	3	US-08-948-564-3	Sequence 3, Appli
45	41	49.4	1831	3	US-09-422-576D-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-568-459A-11
; Sequence 11, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobb Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660

COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israel, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550

```

; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-568-459A-11

Alignment Scores:
Pred. No.: 154 Length: 8220
Score: 47.00 Matches: 9
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 64.23% Mismatches: 3
Query Match: 56.63% Indels: 0
DB: 2 Gaps: 0

US-10-030-937-72 (1-16) x US-08-568-459A-11 (1-8220)

QY 3 LeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
DB 6341 CTTCCAAAAACGATGGAAGTCTTCGGATTAGAAAACCG 6382

RESULT 2
US-08-487-826B-11
; Sequence 11, Application US/08487826B
; Patent No. 5933827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-09-210-288-11

Alignment Scores:
Pred. No.: 154 Length: 8220
Score: 47.00 Matches: 9
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 64.29% Mismatches: 3
Query Match: 56.63% Indels: 0
DB: 3 Gaps: 0
```

```

; ORGANISM: Plasmodium falciparum
US-08-487-826B-11

Alignment Scores:
Pred. No.: 154 Length: 8220
Score: 47.00 Matches: 9
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 64.29% Mismatches: 3
Query Match: 56.63% Indels: 0
DB: 2 Gaps: 0

US-10-030-937-72 (1-16) x US-08-487-826B-11 (1-8220)

QY 3 LeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
DB 6341 CTTCCAAAAACGATGGAAGTCTTCGGATTAGAAAACCG 6382

RESULT 3
US-09-210-288-11
; Sequence 11, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-09-210-288-11

Alignment Scores:
Pred. No.: 154 Length: 8220
Score: 47.00 Matches: 9
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 64.29% Mismatches: 3
Query Match: 56.63% Indels: 0
DB: 3 Gaps: 0
```

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

```
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13

Alignment Scores:
Pred. No.: 447 Length: 19124
Score: 47.00 Matches: 9
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 64.23% Mismatches: 3
Query Match: 56.63% Indels: 0
DB: 2 Gaps: 0

US-10-030-937-72 (1-16) x US-08-487-826B-13 (1-19124)

Qy 3 LeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 13661 CTTCCAAAAGCATGGAAGTCTTCGGATTTAGAAAAGCCG 13702

RESULT 7
US-09-894-844-61/c
; Sequence 61, Application US/09894844
; Patent No. 6696166
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of
; TITLE OF INVENTION: the M. Tuberculosis Complex
; FILE REFERENCE: STAN102CON
; CURRENT APPLICATION NUMBER: US/09/894,844
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/318,191
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/097,936
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 1062
; TYPE: DNA
; ORGANISM: Mycobacteria tuburculosis
; US-09-894-844-61

Alignment Scores:
Pred. No.: 18.1 Length: 1062
Score: 46.00 Matches: 9
Percent Similarity: 78.57% Conservative: 2
```

```
Best Local Similarity: 64.29% Mismatches: 3
Query Match: 55.42% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-72 (1-16) x US-09-894-844-61 (1-1062)

Qy 3 LeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 839 TTGCCGAAGCCCGAGTTCTGAATGCCGAGTTGCCACTGCCA 798

RESULT 8
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
; US-09-103-840A-2

Alignment Scores:
Pred. No.: 5.47e+05 Length: 4403765
Score: 46.00 Matches: 9
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 64.29% Mismatches: 3
Query Match: 55.42% Indels: 0
DB: 3 Gaps: 0

US-10-030-937-72 (1-16) x US-09-103-840A-2 (1-4403765)

Qy 3 LeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 2632604 TTGCCGAAGCCCGAGTTCTGAATGCCGAGTTGCCACTGCCA 2632645

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
```

; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores: 5.47e+05 Length: 4403765
Pred. No.: 46.00 Matches: 8
Score: 68.75% Conservative: 3
Percent Similarity: 50.00% Mismatches: 5
Best Local Similarity: 55.42% Indels: 0
Query Match: 3 Gaps: 0
DB: 0

US-10-030-937-72 (1-16) x US-09-103-840A-2 (1-4403765)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 299240 TACAGCGCACCGAAGCGGCGGCACACCGAGACGACCATCCG 299193

RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores: 5.47e+05 Length: 4411529
Pred. No.: 46.00 Matches: 9
Score: 78.57% Conservative: 2
Percent Similarity: 64.29% Mismatches: 3
Best Local Similarity: 55.42% Indels: 0
Query Match: 3 Gaps: 0
DB: 0

US-10-030-937-72 (1-16) x US-09-103-840A-1 (1-4411529)
QY 3 LeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 2634752 TTGCCGAGCCGAGTTCGTAATGCCGAGCGGATTCGCCACATGCCA 2634793

RESULT 11
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores: 5.47e+05 Length: 4411529
Pred. No.: 46.00 Matches: 8
Score: 68.75% Conservative: 3
Percent Similarity: 50.00% Mismatches: 5
Best Local Similarity: 55.42% Indels: 0
Query Match: 3 Gaps: 0
DB: 0

US-10-030-937-72 (1-16) x US-09-103-840A-1 (1-4411529)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 299129 TACAGCGCACCGAAGCGGCGGCACACCGAGACGACCATCCG 299082

RESULT 12
US-09-543-681A-69
; Sequence 69, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 69
; LENGTH: 924
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-69

Alignment Scores: 36.4 Length: 924
Pred. No.: 44.00 Matches: 7
Score: 78.57% Conservative: 4
Percent Similarity: 50.00% Mismatches: 3
Best Local Similarity: 53.01% Indels: 0
Query Match: 4 Gaps: 0
DB: 0

US-10-030-937-72 (1-16) x US-09-543-681A-69 (1-924)
QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGlu 14
Db 658 TACAGTTTACTCAACCAATTTACTGATCTCGATATTCAA 699

RESULT 13
US-08-911-853-3
; Sequence 3, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,853
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,092
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-911-853-3

Alignment Scores:
Pred. No.: 96.9 Length: 1416
Score: 43.00 Matches: 8
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 51.81% Indels: 0
Gaps: 0

US-10-030-937-72 (1-16) x US-08-911-853-3 (1-1416)
Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 121 TACAGCATTCGACCTTCGACCTGGTGTGCTCAGCGACCTGCGCGCTGCC 168

RESULT 14
US-09-479-409-3
; Sequence 3, Application US/09479409
; Patent No. 6225106
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,409
; FILING DATE:
; PRIOR APPLICATION DATA:
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,409
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-479-409-3

Alignment Scores:
Pred. No.: 96.9 Length: 1416
Score: 43.00 Matches: 8
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 51.81% Indels: 0
Gaps: 0

US-10-030-937-72 (1-16) x US-08-911-853-3 (1-1416)
Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 121 TACAGCATTCGACCTTCGACCTGGTGTGCTCAGCGACCTGCGCGCTGCC 168
```

```
; LENGTH: 1416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-479-409-3

Alignment Scores:
Pred. No.: 96.9 Length: 1416
Score: 43.00 Matches: 8
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 51.81% Indels: 0
Gaps: 0

US-10-030-937-72 (1-16) x US-09-479-409-3 (1-1416)
Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 121 TACAGCATTCGACCTTCGACCTGGTGTGCTCAGCGACCTGCGCGCTGCC 168

RESULT 15
US-09-479-453-3
; Sequence 3, Application US/09479453
; Patent No. 6313283
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,453
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-479-453-3

Alignment Scores:
Pred. No.: 96.9 Length: 1416
Score: 43.00 Matches: 8
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 51.81% Indels: 0
Gaps: 0

US-10-030-937-72 (1-16) x US-09-479-453-3 (1-1416)
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 22:22:21 ; Search time 73.6 Seconds
(without alignments)
1174.701 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 83
Sequence: 1 YSLPKSEFAVPDLPLP 16

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3627888 seqs, 2701811610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US1003037/runat_16112004_153015_3035/app_query.fasta_1.789
-DB=Published Applications NA -QWTF=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US1003037 @CGN 1 1 1223 @runat 16112004 153015_3035
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
- 18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

Result No.	Score	Query Match	Length	DB ID	Description
1	79	95.2	448	11	US-09-969-034-4215
2	79	95.2	1935	10	US-09-971-392-102
3	79	95.2	2384	9	US-09-822-849A-53
4	79	95.2	2436	9	US-09-954-531-380
5	79	95.2	2436	10	US-09-525-978B-81
6	79	95.2	2478	15	US-10-170-385-390
7	79	95.2	250000	15	US-10-225-810-26
8	69	83.1	1983	16	US-10-388-934-167
9	49	59.0	2269	17	US-10-437-963-97626
10	48	57.8	475	17	US-10-021-323-4461
11	48	57.8	475	17	US-10-767-795-2700
12	48	57.8	2796	17	US-10-437-963-97629
13	48	57.8	3335	17	US-10-437-963-97629
14	47	56.6	8220	13	US-10-153-273-11
15	47	56.6	9280	8	US-08-781-986A-131
16	47	56.6	9280	16	US-10-329-624-131
17	47	56.6	68233	15	US-10-034-650-31
18	47	56.6	122614	13	US-10-087-192-1726
19	47	56.6	402850	10	US-09-844-653-5
20	46	55.4	442	18	US-10-425-115-45305
21	46	55.4	535	13	US-10-027-632-246658
22	46	55.4	535	15	US-10-027-632-246658
23	46	55.4	1062	9	US-09-894-844-61
24	46	55.4	1062	16	US-10-388-902-61
25	46	55.4	1062	16	US-10-647-089-61
26	46	55.4	4999	15	US-10-225-486-50
27	45	54.2	1200	16	US-10-282-132A-24975
28	45	54.2	1749	15	US-10-172-118-1397
29	45	54.2	1749	15	US-10-295-027-457
30	45	54.2	1749	15	US-10-295-027-1095
31	45	54.2	1749	16	US-10-342-887-1397
32	45	54.2	1749	18	US-10-422-522-34
33	45	54.2	1764	14	US-10-125-237-20
34	45	54.2	1764	14	US-10-105-891-20
35	45	54.2	2015	14	US-10-176-847-43
36	45	54.2	2282	14	US-10-198-846-13320
37	45	54.2	2282	18	US-10-425-115-68772
38	45	54.2	4146	16	US-10-381-792-3
39	45	54.2	5368	15	US-10-257-021-29
40	45	54.2	5460	16	US-10-287-226-315
41	45	54.2	5472	16	US-10-381-792-4
42	45	54.2	5607	15	US-10-291-172-494
43	45	54.2	5607	16	US-10-221-278-494
44	45	54.2	5884	15	US-10-291-172-118
45	45	54.2	5884	16	US-10-221-278-118

ALIGNMENTS

RESULT 1
US-09-969-034-4215
; Sequence 4215, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwivedi, Poorima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially Expressed in Cancer Tissue
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 4494

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4215
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 241..277, 288, 295, 299, 300, 304, 310, 316, 343, 346, 356,
; LOCATION: 364, 370, 396, 397, 406, 410, 415, 424, 437
; OTHER INFORMATION: n = A,T,C or G
US-09-969-034-4215

Alignment Scores:
Pred. No.: 1.33e-05 Length: 448
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 11 Gaps: 0

US-10-030-937-72 (1-16) x US-09-969-034-4215 (1-448)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 43 TACTCACTGCCCAAGAGCGCAATTGTTGCTGCTGACCTGGAGCTGCC 90

RESULT 2
US-09-971-392-102
; Sequence 102, Application US/09971392
; Publication No. US20030134283A1
; GENERAL INFORMATION:
; APPLICANT: Peterson, David P.
; APPLICANT: Pearson, Cecelia I.
; APPLICANT: Cocks, Benjamin G.
; TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
; FILE REFERENCE: PA-0029 US
; CURRENT APPLICATION NUMBER: US/09/971,392
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 60/237,652
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PERL Program
; SEQ ID NO 102
; LENGTH: 1935
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Template ID: 977615.8
US-09-971-392-102

Alignment Scores:
Pred. No.: 7.16e-05 Length: 1935
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 10 Gaps: 0

US-10-030-937-72 (1-16) x US-09-971-392-102 (1-1935)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 534 TACTCACTGCCCAAGAGCGCAATTGTTGCTGCTGACCTGGAGCTGCC 581

RESULT 3
US-09-822-849A-53
; Sequence 53, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
```

```
; APPLICANT: Agoetino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 2384
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-53

Alignment Scores:
Pred. No.: 9.1e-05 Length: 2384
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-72 (1-16) x US-09-822-849A-53 (1-2384)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 445 TACTCACTGCCCAAGAGCGCAATTGTTGCTGCTGACCTGGAGCTGCC 492

RESULT 4
US-09-954-531-380
; Sequence 380, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 380
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-380

Alignment Scores:
Pred. No.: 9.32e-05 Length: 2436
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 9 Gaps: 0

US-10-030-937-72 (1-16) x US-09-954-531-380 (1-2436)
```

```
QY      1  TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db      491 TACTCACTGCCAAGAGCGAATTGTTGCTGACCTGGAGCTGCC 538

RESULT 5
US-09-525-978B-81
; Sequence 81, Application US/09525978B
; Publication No. US20030049722A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Caras, Ingrid W.
; APPLICANT: Hevezi, Peter
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING MACROPHAGE DEVELOPMENT
; TITLE OF INVENTION: RELATED DISORDERS, COMPOSITIONS, AND METHODS OF
; TITLE OF INVENTION: SCREENING FOR MACROPHAGE DEVELOPMENT MODULATORS
; FILE REFERENCE: A-67413-1/DJB/JDD
; CURRENT APPLICATION NUMBER: US/09/525,978B
; CURRENT FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: USSN 60/124,530
; PRIOR FILING DATE: 1999-03-15
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-525-978B-81

Alignment Scores:
Pred. No.:      9.32e-05      Length:      2436
Score:          79.00      Matches:      15
Percent Similarity: 93.75%      Conservative: 0
Best Local Similarity: 93.75%      Mismatches: 1
Query Match:     95.18%      Indels:      0
DB:              10      Gaps:      0

US-10-030-937-72 (1-16) x US-09-525-978B-81 (1-2436)
QY      1  TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db      491 TACTCACTGCCAAGAGCGAATTGTTGCTGACCTGGAGCTGCC 538

RESULT 6
US-10-170-385-390
; Sequence 390, Application US/10170385
; Publication No. US2003020372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 390
; LENGTH: 2478
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-170-385-390

Alignment Scores:
Pred. No.:      9.51e-05      Length:      2478
Score:          79.00      Matches:      15
Percent Similarity: 93.75%      Conservative: 0
Best Local Similarity: 93.75%      Mismatches: 1
Query Match:     95.18%      Indels:      0
DB:              15      Gaps:      0

US-10-030-937-72 (1-16) x US-10-170-385-390 (1-2478)
QY      1  TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db      528 TACTCACTGCCAAGAGCGAATTGTTGCTGACCTGGAGCTGCC 575

RESULT 7
US-10-225-810-26/c
; Sequence 26, Application US/10225810
; Publication No. US20030157512A1
; GENERAL INFORMATION:
; APPLICANT: Birmingham, Jr., John R.
; TITLE OF INVENTION: Tramdorins and Methods of Using Tramdorin
; FILE REFERENCE: McLaugh-07165
; CURRENT APPLICATION NUMBER: US/10/225,810
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
; LENGTH: 250000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (33774)..(33774)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (42953)..(43052)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45557)..(45656)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (48203)..(48302)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (49551)..(49650)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51561)..(51660)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (52722)..(52821)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (53864)..(53963)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55290)..(55389)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (56674)..(56773)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (57879)..(57978)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (78952)..(79051)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (85316)..(85415)
; OTHER INFORMATION: n is a, c, g, or t
US-10-225-810-26
```

```
Alignment Scores:
Pred. No.: 0.019 Length: 250000
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 15 Gaps: 0
```

US-10-030-937-72 (1-16) x US-10-225-810-26 (1-250000)

```
QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 249623 TACTCACTGCCAGAGCGCAATTCGTTGTGCTGACCTGGAGCTGCC 249576
```

RESULT 8

```
US-10-388-934-167
; Sequence 167, Application US/10388934
; Publication No. US20040005547A1
; GENERAL INFORMATION:
; APPLICANT: Boess, Franziska
; APPLICANT: Suter-Dick, Laura
; APPLICANT: Wolf, Detlef
; TITLE OF INVENTION: BIOMARKERS AND EXPRESSION PROFILES FOR TOXICOLOGY
; FILE REFERENCE: 21199
; CURRENT APPLICATION NUMBER: US/10/388,934
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 02005336.9
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 02015657.6
; PRIOR FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 862
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 167
; LENGTH: 1983
; TYPE: DNA
; ORGANISM: Rattus sp.
US-10-388-934-167
```

```
Alignment Scores:
Pred. No.: 0.00568 Length: 1983
Score: 69.00 Matches: 13
Percent Similarity: 81.25% Conservative: 0
Best Local Similarity: 81.25% Mismatches: 3
Query Match: 83.13% Indels: 0
DB: 16 Gaps: 0
```

US-10-030-937-72 (1-16) x US-10-388-934-167 (1-1983)

```
QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
Db 462 TACTCACTGCCTCGAGCAACTTCACAGTGCCTGATCTGGAGCTTCCA 509
```

RESULT 9

```
US-10-437-963-4335
; Sequence 4335, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
```

```
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; OTHER INFORMATION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 4335
; LENGTH: 2269
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(2269)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MBT4530_11228C.1
US-10-437-963-4335
```

```
Alignment Scores:
Pred. No.: 39.4 Length: 2269
Score: 49.00 Matches: 8
Percent Similarity: 92.31% Conservative: 4
Best Local Similarity: 61.54% Mismatches: 1
Query Match: 59.04% Indels: 0
DB: 17 Gaps: 0
```

US-10-030-937-72 (1-16) x US-10-437-963-4335 (1-2269)

```
QY 3 LeuProLysSerGluPheAlaValProAspLeuGluLeu 15
Db 221 GTTCACGAAGTGAATTTCCATACCTGATTGGAGCTT 259
```

RESULT 10

```
US-10-021-323-4461/c
; Sequence 4461, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 4461
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3825-025-Q6-N6-C6
US-10-021-323-4461
```

```
Alignment Scores:
Pred. No.: 10.1 Length: 475
Score: 48.00 Matches: 10
Percent Similarity: 73.33% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 4
Query Match: 57.83% Indels: 0
DB: 17 Gaps: 0
```

US-10-030-937-72 (1-16) x US-10-021-323-4461 (1-475)

```
QY 2 SerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
|||||
:::|||||
```

Db 435 TCCTCCCAAGGACGATGCTGTCCTCCATTCCTGGAACTACCA 391

RESULT 11
US-10-767-795-2700
; Sequence 2700, Application US/10767795
; Publication No. US20040181830A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10/767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 2700
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C49222_1
US-10-767-795-2700

Alignment Scores:
Pred. No.: 10.1 Length: 475
Score: 48.00 Matches: 10
Percent Similarity: 73.33% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 4
Query Match: 57.83% Indels: 0
DB: 17 Gaps: 0

US-10-030-937-72 (1-16) x US-10-767-795-2700 (1-475)

QY 2 SerLeuProlySerGluPheAlaValProAspLeuGluLeuPro 16
Db 41 TCCTCCCAAGGACGATGCTGTCCTCCATTCCTGGAACTACCA 85

RESULT 12
US-10-437-963-97626/c
; Sequence 97626, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 97626
; LENGTH: 2796
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_95610C.1
US-10-437-963-97626

Alignment Scores:
Pred. No.: 77.3 Length: 2796
Score: 48.00 Matches: 9
Percent Similarity: 80.00% Conservative: 3
Best Local Similarity: 60.00% Mismatches: 3
Query Match: 57.83% Indels: 0
DB: 17 Gaps: 0

US-10-030-937-72 (1-16) x US-10-437-963-97626 (1-2796)

QY 1 TyrSerLeuProlySerGluPheAlaValProAspLeuGluLeu 15
Db 808 TTTATTTCGCCAGAAAGTTCCTTCGCCATTCGTTGATCTTGAGTTG 764

RESULT 13
US-10-437-963-97629/c
; Sequence 97629, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 97629
; LENGTH: 3335
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_95613C.1
US-10-437-963-97629

Alignment Scores:
Pred. No.: 94.6 Length: 3335
Score: 48.00 Matches: 9
Percent Similarity: 80.00% Conservative: 3
Best Local Similarity: 60.00% Mismatches: 3
Query Match: 57.83% Indels: 0
DB: 17 Gaps: 0

US-10-030-937-72 (1-16) x US-10-437-963-97629 (1-3335)

QY 1 TyrSerLeuProlySerGluPheAlaValProAspLeuGluLeu 15
Db 1348 TTTATTTCGCCAGAAAGTTCCTTCGCCATTCGTTGATCTTGAGTTG 1304

RESULT 14
US-10-153-273-11
; Sequence 11, Application US/10153273
; Publication No. US20020169305A1
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Welles, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/153,273
; FILING DATE: 21-May-2002
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-153-273-11

```

```

Alignment Scores:
Pred. No.: 412 Length: 8220
Score: 47.00 Matches: 9
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 64.29% Mismatches: 3
Query Match: 56.63% Indels: 0
DB: 13 Gaps: 0

```

```

US-10-030-937-72 (1-16) x US-10-153-273-11 (1-8220)
QY 3 LeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
DB 6341 CTTCCAAAACGATGGAGACTGTTCCGGATTAGAAAAGCCG 6382

```

```

RESULT 15
US-08-781-986A-131
; Sequence 131, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:

```

```

;
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 131:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-131
Alignment Scores:
Pred. No.: 473 Length: 9280
Score: 47.00 Matches: 8
Percent Similarity: 86.67% Conservative: 5
Best Local Similarity: 53.33% Mismatches: 2
Query Match: 56.63% Indels: 0
DB: 8 Gaps: 0
US-10-030-937-72 (1-16) x US-08-781-986A-131 (1-9280)
QY 2 SerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
DB 6399 ACGCTTCGAAATCAAGTTTGCTTACCACCAATATTACATACCA 6443

```

Search completed: November 19, 2004, 03:29:55
Job time : 87.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 18, 2004, 19:16:51 ; Search time 514.773 Seconds
(without alignments)
1132.607 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 83

Sequence: 1 YSLPKSEFAVDPDLEP 16

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPTO.spool.p/US10030937/runat.16112004.153014.2925/app.query.fasta.1.789
-DB=EST -OFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030937 @CGN_1.1.9321 @runat.16112004.153014.2925 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: EST:

1: gb_est1:*
2: gb_est2:*
3: gb_hc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gse1:*
9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83	100.0	784	4 BG323734	602421833
2	83	100.0	885	6 CA453926	AGENCOURT
3	79	95.2	360	1 AA224748	nc1202.s
4	79	95.2	366	2 BE182886	BE182886
5	79	95.2	377	1 AA224768	AA224768
6	79	95.2	388	2 BF916278	BF916278
7	79	95.2	389	2 BF949518	BF949518
8	79	95.2	554	4 BG830178	602764754
9	79	95.2	559	1 AU280628	AU280628

10	79	95.2	601	5 BX506263	BX506263
11	79	95.2	611	4 BG623044	602647926
12	79	95.2	643	1 AU135438	AU135438
13	79	95.2	646	6 CD703135	EST19726
14	79	95.2	651	4 BM723945	UI-E-B01-
15	79	95.2	658	2 BF509172	BF509172
16	79	95.2	711	4 BG479322	602526351
17	79	95.2	712	6 CA422926	CA422926
18	79	95.2	726	4 BG762599	602734472
19	79	95.2	733	4 BI856212	603382936
20	79	95.2	760	4 BG770447	602734356
21	79	95.2	784	4 BI838554	603086219
22	79	95.2	807	4 BI086562	602849951
23	79	95.2	813	4 BG913328	602812047
24	79	95.2	818	1 AL552056	AL552056
25	79	95.2	842	4 BI091220	602856051
26	79	95.2	847	5 BQ220522	AGENCOURT
27	79	95.2	858	4 BI820051	603037236
28	79	95.2	863	4 BG478588	602524087
29	79	95.2	895	5 BQ676659	AGENCOURT
30	79	95.2	909	5 BQ643369	AGENCOURT
31	79	95.2	910	1 AL560604	AL560604
32	79	95.2	912	1 AL543858	AL543858
33	79	95.2	937	5 BUI151364	AGENCOURT
34	79	95.2	948	5 BUI56606	AGENCOURT
35	79	95.2	949	5 BQ600662	AGENCOURT
36	79	95.2	973	4 BM474816	AGENCOURT
37	79	95.2	986	5 BQ880758	AGENCOURT
38	79	95.2	994	1 AL548441	AL548441
39	79	95.2	997	4 BM561693	AGENCOURT
40	79	95.2	1018	2 BE735010	601567832
41	79	95.2	1031	4 BM476220	AGENCOURT
42	79	95.2	1033	5 BQ662907	AGENCOURT
43	79	95.2	1060	1 AL550565	AL550565
44	79	95.2	1067	2 BF528447	602043611
45	79	95.2	1087	4 BM805102	AGENCOURT

ALIGNMENTS

RESULT 1
BG323734
LOCUS 602421833F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4559888 5',
mRNA sequence. 784 bp mRNA linear EST 27-FEB-2001

ACCESSION BG323734

VERSION BG323734.1 GI:13130171

KEYWORDS EST.

SOURCE Homo sapiens

ORGANISM Homo sapiens (human)

REFERENCE NIH-MGC http://mgc.nci.nih.gov/

AUTHORS 1 (bases 1 to 784)

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DP

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1CM1268 row: a column: 09

High quality sequence stop: 777.

Location/Qualifiers

1. 784

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4559888"

FEATURES

source

/tissue type="renal cell adenocarcinoma"
 /lab host="DH10B (phage-resistant)"
 /clone lib="NIH MGC 14"
 /note="Organ: kidney; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.: 0.000454 Length: 784
 Score: 83.00 Matches: 16
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-030-937-72 (1-16) x BG323734 (1-784)

Qy 1 TyrSerLeuProLySerGluPheAlaValProAspLeuGluLeuPro 16
 Db 438 TACTCAGTCCCAAGAGCGAATTCGCTGTGCTGACCTGGAGCTGCC 485

RESULT 2
 CA453926
 LOCUS
 DEFINITION AGENCOURT_10738491 MAPcL Homo sapiens cDNA clone IMAGE:6718573 5',
 mRNA sequence.
 ACCESSION
 VERSION CA453926
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 (bases 1 to 885)
 NIH-MGC http://mgc.nci.nih.gov/.
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Kristi A. Eglund, Ira Pastan
 cDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM14275 row: b column: 13
 High quality sequence stop: 427.

FEATURES

source

1..885
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6718573"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
 HRT-1, LNCaP"
 /lab_host="EMD10B"
 /clone_lib="MAPcL"
 /note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
 Subtracted with brain, liver, lung, kidney and muscle.
 Directionally cloned. Priming method: oligo-dT. Average
 insert size: 1800 bp. Library amplification: 26,000 fold.
 Kristi A. Eglund, James J. Vincent, Robert Strausberg,
 Bungkook Lee & Ira Pastan. Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

ORIGIN

Alignment Scores:

Pred. No.: 0.0009 Length: 360
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0

Alignment Scores:

Pred. No.: 0.000525 Length: 885
 Score: 83.00 Matches: 16
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-030-937-72 (1-16) x CA453926 (1-885)

Qy 1 TyrSerLeuProLySerGluPheAlaValProAspLeuGluLeuPro 16
 Db 398 TACTCAGTCCCAAGAGCGAATTCGCTGTGCTGACCTGGAGCTGCC 445

RESULT 3
 AA224748/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 (bases 1 to 360)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
 M.D., Michael Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: David B. Krizman, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 669 Std Error: 0.00
 Seq primer: -41m3 fwd. Et from Amersham
 High quality sequence stop: 298.

FEATURES

source

1..360
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1007906"
 /sex="Male"
 /dev stage="45 years old"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Pr1"
 /note="Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st
 strand cDNA was primed with oligo(dT)17 on 50 ng of
 DNase-treated, total cellular RNA obtained from
 5,000-10,000 microdissected, histologically normal
 prostate epithelial cells. Double-stranded cDNA was
 ligated to EcoRI adaptors, 5 cycles of PCR applied to the
 cDNA with an adaptor-specific primer, and the resulting
 PCR product subcloned into pAMP10 by the UDG-cloning
 method (Life Technologies). Average insert size is 600
 bp. NOTE: Not directionally cloned. This library was
 constructed by David Krizman."

Best Local Similarity: 93.75%
 Query Match: 95.18%
 DB: 1

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
 Db 209 TACTCACTGCCAAGAGCGAATTGTTGCTGCTGACCTGGAGCTGCC 256

US-10-030-937-72 (1-16) x AA224748 (1-360)

RESULT 4
 BE182886
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT

BE182886 366 bp mRNA linear EST 22-JUN-2000
 CM4-HT0652-150400-143-C03 HT0652 Homo sapiens cDNA, mRNA sequence.
 BE182886
 BE182886.1 GI:8662062
 EST.
 Homo sapiens (human)
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 10737800
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM4-HT0652-150
 400-143-C03&t3=2000-04-15&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 366.
 Location/Qualifiers

1. 366
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="HT0652"

/note="Organ: head neck; Vector: puc18; Site 1: SmaI;
 Site 2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

ORIGIN

Alignment Scores:
 Pred. No.: 0.000918 Length: 366
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservatives: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 2 Gaps: 0

US-10-030-937-72 (1-16) x BE182886 (1-366)

RESULT 5
 AA224768
 LOCUS
 DEFINITION

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

AA224768 377 bp mRNA linear EST 19-AUG-1997
 nc1209.r1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:1007921
 similar to gb:X62078 GANGLIOSIDE GM2 ACTIVATOR PRECURSOR (HUMAN);
 mRNA sequence.
 AA224768
 AA224768.1 GI:1846077
 EST.
 Homo sapiens (human)
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
 M.D., Michael Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
 DNA sequencing by: Washington University Genome Sequencing Center
 Clones distributed by: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: -28ml3 rev1 ET from Amersham
 High quality sequence stop: 167.
 Location/Qualifiers

1. 377
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1007921"
 /sex="Male"
 /dev_stage="45 years old"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Prl"
 /note="Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st
 strand cDNA was primed with oligo(dT)17 on 50 ng of
 DNase-treated, total cellular RNA obtained from
 5,000-10,000 microdissected, histologically normal
 prostate epithelial cells. Double-stranded cDNA was
 ligated to EcoRI adaptors, 5 cycles of PCR applied to the
 cDNA with an adaptor-specific primer, and the resulting
 PCR product subcloned into pAMP10 by the UDG-cloning
 method (Life Technologies). Average insert size is 600
 bp. NOTE: Not directionally cloned. This library was
 constructed by David Krizman."

ORIGIN

Alignment Scores:
 Pred. No.: 0.000952 Length: 377
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservatives: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 1 Gaps: 0

US-10-030-937-72 (1-16) x AA224768 (1-377)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
 Db 174 TACTCACTGCCAAGAGCGAATTGTTGCTGCTGACCTGGAGCTGCC 221

RESULT 6
 BF916278

LOCUS BF916278 388 bp mRNA linear EST 18-JAN-2001
 DEFINITION PM1-UT0104-111200-004-b12 UT0104 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF916278
 VERSION BF916278.1 GI:12307736
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 388)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?ci=PM1-UT0104-111200-004-b12&t3=2000-12-11&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 7
 High quality sequence stop: 388.

FEATURES
 source
 1..388
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="UT0104"
 /note="Organ: uterus_tumor; Vector: puc18; Site: 1: Smal; Site 2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
 Alignment Scores:
 Pred. No.: 0.000985 Length: 388
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 2 Gaps: 0
 US-10-030-937-72 (1-16) x BF916278 (1-388)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
 |||
 DB 248 TACTCACTGCCAAGAGCGAATTCTGTGCTGACCTGGAGCTGCC 295
 |||
 RESULT 7
 BF949518 389 bp mRNA linear EST 22-JAN-2001
 LOCUS BF949518
 DEFINITION MR3-NN0218-031100-002-d08 NN0218 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF949518
 VERSION BF949518.1 GI:12366793
 KEYWORDS EST.

SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 389)
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT

CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?ci=MR3-NN0218-031100-002-d08&t3=2000-11-03&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 389.

FEATURES
 source
 1..389
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="NN0218"
 /note="Organ: nervous normal; Vector: puc18; Site: 1: Smal; Site 2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
 Alignment Scores:
 Pred. No.: 0.000988 Length: 389
 Score: 79.00 Matches: 15
 Percent Similarity: 93.75% Conservative: 0
 Best Local Similarity: 93.75% Mismatches: 1
 Query Match: 95.18% Indels: 0
 DB: 2 Gaps: 0
 US-10-030-937-72 (1-16) x BF949518 (1-389)

QY 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16
 |||
 DB 223 TACTCACTGCCAAGAGCGAATTCTGTGCTGACCTGGAGCTGCC 270
 |||
 RESULT 8
 BF830178 554 bp mRNA linear EST 22-MAY-2001
 LOCUS BF830178
 DEFINITION 602764754F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4906909 5', mRNA sequence.
 ACCESSION BF830178
 VERSION BF830178.1 GI:14177765
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 554)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-k@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCMI810 row: 1 column: 14
 High quality sequence stop: 554.

FEATURES
 source
 1..554
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4906909"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_42"
 /notes="Organ: pancreas; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library. |"

ORIGIN

Alignment Scores:

Pred. No.:	0.00151	Length:	554
Score:	79.00	Matches:	15
Percent Similarity:	93.75%	Conservative:	0
Best Local Similarity:	93.75%	Mismatches:	1
Query Match:	95.18%	Indels:	0
DB:	4	Gaps:	0

US-10-030-937-72 (1-16) x BG830178 (1-554)

Qy 1 Ty-SerLeuProlySerGluPheAlaValProAspLeuGluLeuPro 16
Db 506 TACTCACTGCCAAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCCA 553

RESULT 9
AU280628
LOCUS AU280628 NCRRM1 Homo sapiens cDNA clone NCRRM1000016 5', mRNA EST 31-JUL-2003
DEFINITION sequence.
ACCESSION AU280628
VERSION AU280628.1 GI:28299855
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 559)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Inabayashi, H., Mori, T., Gojo, S., Kiyono, T., Sugiyama, T., Irie, R., Isogai, T., Hata, J., Tomoya, Y. and Umezawa, A.
 Redifferentiation of dedifferentiated chondrocytes and chondrogenesis of human bone marrow stromal cells via chondrosphere formation with expression profiling by large-scale cDNA analysis
JOURNAL Exp. Cell Res. 288 (1), 35-50 (2003)
MEDLINE 22760698
PUBMED 12878157
COMMENT Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human cDNA Project, Sugiyama, T.; Wakamatsu, A.; Irie, R.; Umezawa, A.; Fukuma, M.; Kusakari, S.; Hata, J.; Ishii, S.; Yamamoto, J.; Isono, Y.; Saito, K.; Nakamura, Y.; Masuho, Y.; Nagai, K.; Isogai, T.
 HRI human cDNA project; cDNA library construction & 5'-end one pass sequencing; Helix Research Institute.

FEATURES

source
 1..559
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NCRRM1000016"
 /cell_type="embryonal carcinoma"
 /clone_lib="NCRRM1"
 /notes="Vector: pME18SPL3; mRNA from uninduced embryonal carcinoma"

ORIGIN

Alignment Scores:

Pred. No.:	0.00153	Length:	559
Score:	79.00	Matches:	15
Percent Similarity:	93.75%	Conservative:	0
Best Local Similarity:	93.75%	Mismatches:	1
Query Match:	95.18%	Indels:	0
DB:	1	Gaps:	0

US-10-030-937-72 (1-16) x AU280628 (1-559)

Qy 1 Ty-SerLeuProlySerGluPheAlaValProAspLeuGluLeuPro 16
Db 465 TACTCACTGCCAAGAGCGAATTCGTTGCTGCTGACCTGGAGCTGCC 512

RESULT 10

BX506263
LOCUS BX506263 601 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFP686F03234_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone DKFP686F03234 5', mRNA sequence.
ACCESSION BX506263
VERSION BX506263.1 GI:32037321
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 601)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Ottenwaelder, B., Obermaier, B., Deutschbauer, S., Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.
 EST (Ottenwaelder, B., Obermaier, B., Deutschbauer, S., Mewes, H.W., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS
 MIPS
 Ingolstaedter Landstr.1, D-95764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by MediGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No sl sequence available.
 This clone (DKFP686F03234) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source
 1..601
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFP686F03234"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="586 (synonym: hlcc3)"

AUTHORS Liu, X.-Q., Zhou, Y., Zhang, L.-J., Xu, H., Chen, H.-K., Pan, Z.-G. and Zeng, Y.-X.

TITLE Transcriptional Gene Expression Profile of Human Nasopharynx

JOURNAL Unpublished (2003)

COMMENT Contact: YiXin Zeng

Cancer Center

Sun Yat-sen University

651 Dongfeng Road East, Guangzhou 510060, China

Tel: 86-1380-9770-743

Fax: 86-20-8775-4506

Email: yxzeng@gzsums.edu.cn.

FEATURES Location/Qualifiers

source

1. 646

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="normal nasopharynx"

/clone_lib="human nasopharynx"

/note="ESTs generated from a normal nasopharynx cDNA

library from southern Chinese"

ORIGIN

Alignment Scores:
Pred. No.: 0.00181 Length: 646
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 6 Gaps: 0

US-10-030-937-72 (1-16) x CD703135 (1-646)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16

Db 495 TACTCACTGCCCAAGAGCGAATCGTTGCTGCTGACCTGGAGTGC 542

RESULT 14

BM723945

LOCUS

DEFINITION UI-E-E01-aix-1-18-0-UI.r1 UI-E-E01 Homo sapiens cDNA clone

UI-E-E01-aix-1-18-0-UI 5', mRNA sequence.

ACCESSION BM723945

VERSION BM723945.1 GI:19045276

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL 97044477

MEDLINE 8889548

PUBMED

COMMENT Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

source

1. 651

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-E-E01-aix-1-18-0-UI"

/tissue_type="fetal eye"

/dev_stage="fetal"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-E-E01"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-E-E01 is a normalized cDNA library containing the

following tissue(s): fetal eye. The library was

constructed according to Bonaldo, Lennon and Soares;

Genome Research, 6:791-806, 1996. First strand cDNA

synthesis was primed with an oligo-dT primer containing a

Not I site. Double stranded cDNA was ligated to an EcoR I

adaptor, digested with Not I, and cloned directionally

into pT73-Pac vector. The oligonucleotide used to prime

the synthesis of first-strand cDNA contains a library tag

sequence that is located between the Not I site and the

(dT)18 tail. The sequence tag for this library is

CGCGTATACC. This library was created for the program, Gene

Discovery in the Visual System, supported by National Eye

Institute (NEI)."

ORIGIN

Alignment Scores:
Pred. No.: 0.00183 Length: 651
Score: 79.00 Matches: 15
Percent Similarity: 93.75% Conservative: 0
Best Local Similarity: 93.75% Mismatches: 1
Query Match: 95.18% Indels: 0
DB: 4 Gaps: 0

US-10-030-937-72 (1-16) x BM723945 (1-651)

Qy 1 TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro 16

Db 437 TACTCACTGCCCAAGAGCGAATCGTTGCTGCTGACCTGGAGTGC 484

RESULT 15

BF509172

LOCUS

DEFINITION BF509172 698 bp mRNA linear EST 06-DEC-2000

IMAGE:3086203 3', mRNA sequence.

ACCESSION BF509172

VERSION BF509172.1 GI:11592470

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Oligo-dT track not found, Not I site shown in beginning of sequence

is likely internal to the message. cDNA Library Preparation: M.B.

Soares Lab Clone distribution: NCI-CGAP clone distribution

information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA=No.

Location/Qualifiers

1. 698

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3086203"

/lab_host="DH10B (Life Technologies)"

/clone_lib="NCI-CGAP Sub8"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not 1; Site 2: Eco RI; NCI CGAP Sub8 is a subtracted library derived from NCI CGAP Sub5. The NCI CGAP Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI CGAP Sub5 was used as a tracer in a subtractive hybridization with a driver comprising a pool of clones from NCI CGAP Sub5 (IMAGE clone Ids 2732833-2737415, 3068040-3069191; 25% of the driver population), a pool of clones from NCI CGAP Sub4 (IMAGE clone Ids 2723592-2729326; 25% of the driver population), NCI CGAP Sub6 (pool AIF-AJU, IMAGE Ids 2728969-2733190; 25% of the driver population), and NCI CGAP Sub7 (IMAGE Ids 3069192-3072238, 3081864-3084550; 25% of the driver population). Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.

TAG_SEQ=None found"

ORIGIN

Alignment Scores:			
Pred. No.:	0.00199	Length:	698
Score:	79.00	Matches:	15
Percent Similarity:	93.75%	Conservative:	0
Best Local Similarity:	93.75%	Mismatches:	1
Query Match:	95.18%	Indels:	0
DB:	2	Gaps:	0

US-10-030-937-72 (1-16) x BF509172 (1-698)

QY	1	TyrSerLeuProLysSerGluPheAlaValProAspLeuGluLeuPro	16
Db	482	TACTCACTGCCCAAGAGCGAATTCGTTGTCCTGACCTGGAGCTGCC	529

Search completed: November 19, 2004, 00:22:26
Job time : 516.773 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 06:12:35 ; Search time 4.65196 Seconds
(without alignments)
1233.819 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 83

Sequence: 1 YSLPKSEFAVPDLELP 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83	100.0	16	AAB31927	Aab31927 Amino aci
2	83	100.0	193	AAB31897	Aab31897 Amino aci
3	79	95.2	178	AAB31898	Aab31898 Amino aci
4	79	95.2	178	ABG31346	ABg31346 Non-glyco
5	79	95.2	189	AAB31900	Aab31900 Amino aci
6	79	95.2	193	AAB31901	Aab31901 Amino aci
7	79	95.2	193	AAB31902	Aab31902 Amino aci
8	79	95.2	193	AAB31904	Aab31904 Amino aci
9	79	95.2	193	AAB31896	Aab31896 Amino aci
10	79	95.2	193	AAB31928	Aab31928 Amino aci
11	79	95.2	193	AAB31903	Aab31903 Amino aci
12	79	95.2	193	ABG00720	ABg00720 Novel hum
13	79	95.2	193	ABG31345	ABg31345 Human GM2
14	79	95.2	193	APP65212	App65212 Hypoxia-r
15	79	95.2	193	ADN95858	Adn95858 Human BEC
16	79	95.2	193	ADN03620	Adn03620 Antipsori
17	79	95.2	193	ADQ17712	Adq17712 Human sof
18	79	95.2	200	AAB31899	Aab31899 Amino aci
19	69	83.1	199	AAW10656	Aaw10656 Rat GM2 a
20	47	56.6	2703	AAW70236	Aaw70236 P. falcip
21	47	56.6	2710	AAW22482	Aaw22482 Plasmodiu
22	47	56.6	2710	AAW77904	Aaw77904 P. falcip
23	47	56.6	3060	AAW22475	Aaw22475 Plasmodiu
24	47	56.6	3060	AAW77905	Aaw77905 Plasmodiu
25	44	53.0	307	ADF03956	Adf03956 Bacterial

26	44	53.0	329	7	ADB70066	C. neofor
27	43.5	52.4	1383	5	ABG69027	Abg69027 Rat peria
28	43.5	52.4	1383	5	ABG69031	Abg69031 Rat peria
29	43.5	52.4	1383	5	ABG69032	Abg69032 Rat peria
30	43.5	52.4	1383	8	ADF08475	Adf08475 Rat peria
31	43.5	52.4	1389	5	ABG69024	Abg69024 Rat peria
32	43	51.8	471	2	AAW53825	Aaw53825 Pseudomon
33	43	51.8	471	3	AAW82589	Aaw82589 Pseudomon
34	43	51.8	471	4	AAW82254	Aaw82254 Pseudomon
35	43	51.8	471	5	AAW13655	Aaw13655 Pseudomon
36	43	51.8	812	7	ABO67765	AbO67765 Klebsiell
37	43	51.8	828	6	ABU32220	Abu32220 Protein e
38	43	51.8	937	5	ABP30468	Abp30468 Streptoco
39	43	51.8	948	5	ABP27191	Abp27191 Streptoco
40	42.5	51.2	1391	5	ABG69030	Abg69030 Mouse per
41	42.5	51.2	1398	7	ADJ69201	Adj69201 Human hea
42	42.5	51.2	1461	5	ABG60913	Abg60913 Protein e
43	42.5	51.2	1461	5	AAU84363	Aau84363 Novel hum
44	42.5	51.2	1461	5	ABG69033	Abg69033 Human per
45	42.5	51.2	1461	5	AAO19407	Aao19407 Human mol

ALIGNMENTS

RESULT 1
AAB31927
ID AAB31927 standard; peptide; 16 AA.

XX AC AAB31927;

XX DT 15-MAY-2001 (first entry)

XX DE Amino acid sequence of a peptide fragment of a human protein.

XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX OS Homo sapiens.

XX PN WO200105422-A2.

XX PD 25-JAN-2001.

XX PF 17-JUL-2000; 2000WO-FR002057.

XX PR 15-JUL-1999; 99FR-00009372.

XX PA (INMR) BIOMERIEUX STELHYS.

XX PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX XX WPI; 2001-159475/16.

XX DR Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand.

XX XX Claim 24; Page 108; 209pp; French.

XX CC The present sequence represents a peptide fragment of a human protein,
XX which is used in the method of the invention. The specification describes
XX a method which uses at least one polypeptide or polynucleotide sequence
XX belonging to the perlecan, precursor of the retinol-binding plasma
XX protein, precursor of the ganglioside GM2 activator, calgranulin B or
XX saposin B protein families. The method is used for detecting, preventing
XX or treating a degenerative, neurological and/or auto-immune disease. The
XX polynucleotides and polypeptides are used for diagnosis, prognosis,
XX prevention and treatment of multiple sclerosis (in its various forms and
XX phases). They may also be useful in cases of e.g. Alzheimer's and

CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 16 AA;

Query Match 100.0%; Score 83; DB 4; Length 16;
 Best Local Similarity 100.0%; Pred. No. 3.5e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
 |||||
 DB 1 YSLPKSEFAVPDLELP 16

RESULT 2

AAB31897
 ID AAB31897 standard; protein; 193 AA.

XX AAB31897;

AC AAB31897;

DT 15-MAY-2001 (first entry)

XX Amino acid sequence of a mutant ganglioside GM2 activator protein.

DE Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX Homo sapiens.

OS WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INMR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

PS Claim 25; Page 159-160; 209pp; French.

XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 193 AA;

Query Match 100.0%; Score 83; DB 4; Length 193;
 Best Local Similarity 100.0%; Pred. No. 7e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
 |||||
 DB 145 YSLPKSEFAVPDLELP 160

RESULT 3

AAB31898

ID AAB31898 standard; protein; 178 AA.

XX AAB31898;

XX 15-MAY-2001 (first entry)

XX Amino acid sequence of a human protein.

DE Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX Homo sapiens.

XX WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INMR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.

XX Claim 1; Page 160; 209pp; French.

XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells

XX Sequence 178 AA;

Query Match 95.2%; Score 79; DB 4; Length 178;
 Best Local Similarity 93.8%; Pred. No. 3.1e-05;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
 |||||
 DB 130 YSLPKSEFAVPDLELP 145

RESULT 4
 ABG31346
 ID ABG31346 standard; protein; 178 AA.
 XX
 AC ABG31346;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Non-glycosylated human GM2 activator protein containing His6-tag.
 XX
 KW Human; GM2 activator protein; ganglioside; platelet activating factor;
 KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
 KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
 KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
 KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
 KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
 KW antiinflammatory; antiasthmatic; leukocyte accumulation; virucide;
 KW antiallergic; hepatotropic; antiarthritic; antirheumatic;
 KW tranquilizer.
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Binding-site 5..10
 FT /label= His6 tag
 FT /note= "Nickel chelating region used for purification of
 FT the protein"
 FT Cleavage-site 11..16
 FT /label= Factor_X_cleavage_site
 FT
 XX US6423680-B1.
 XX
 XX 23-JUL-2002.
 XX
 XX 30-OCT-1998; 98US-00183841.
 XX
 XX 30-OCT-1998; 98US-00183841.
 XX
 XX (HSCR-) HSC RES & DEV LP.
 XX
 XX Rigat B, Reynaud D, Mahuran D;
 XX
 XX WPI; 2002-664636/71.
 XX
 XX Composition useful for treating inflammatory conditions e.g. asthma
 XX comprises GM2 activator protein or GM2 activator peptide in combination
 XX with a carrier.
 XX
 XX Example 1; Fig 2; 11pp; English.
 XX
 XX The present invention relates to a composition comprising GM2
 XX (ganglioside) activator protein or a GM2 activator peptide derived from
 XX the GM2 activator protein in combination with a carrier. The composition
 XX comprises the protein or peptide in an amount of 1-100 mg. The GM2
 XX activator protein is capable of inhibiting platelet activating factor
 XX (PAF). The composition of the invention is useful for treating
 XX inflammatory disorders e.g. inflammatory bowel disease, asthma,
 XX autoimmune disease (such as lupus), hypersensitivity infection,
 XX rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
 XX gout and tissue-specific conditions (e.g. glomerulonephritis and
 XX hepatitis). The composition is capable of inhibiting platelet activating
 XX factor, is non-toxic, is efficacious and presents less severe side
 XX effects, including redness, swelling, pain and polymorphonuclear
 XX leukocyte accumulation at the inflammatory site and other associated
 XX cellular responses. The present sequence represents a non-glycosylated
 XX human GM2 activator protein prepared using a His6-tag bacterial
 XX expression system
 XX
 XX Sequence 178 AA;
 XX
 XX Query Match 95.2%; Score 79; DB 5; Length 178;

Best Local Similarity 93.8%; Pred. No. 3.1e-05;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YSLPKSEFAVPDLELP 16
 |||||
 Db 130 YSLPKSEFVVPDLELP 145
 |||||
 RESULT 5
 AAB31900
 ID AAB31900 standard; protein; 189 AA.
 XX
 AC AAB31900;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of a human protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 XX WO200105422-A2.
 XX
 XX 25-JAN-2001.
 XX
 XX 17-JUL-2000; 2000WO-FR002057.
 XX
 XX 15-JUL-1999; 99FR-00009372.
 XX
 XX (INMR) BIOMERIEUX STELHYS.
 XX
 XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX
 XX WPI; 2001-159475/16.
 XX
 XX Detecting, preventing and treating degenerative, neurological and
 XX autoimmune diseases, particularly multiple sclerosis, using specified
 XX polypeptides or related nucleic acid or ligand.
 XX
 XX Claim 1; Page 161-162; 209pp; French.
 XX
 XX The present sequence represents a human protein, which is used in the
 XX method of the invention. The specification describes a method which uses
 XX at least one polypeptide or polynucleotide sequence belonging to the
 XX perlecan, precursor of the retinol-binding plasma protein, precursor of
 XX the ganglioside GM2 activator, calgranulin B or saposin B protein
 XX families. The method is used for detecting, preventing or treating a
 XX degenerative, neurological and/or auto-immune disease. The
 XX polynucleotides and polypeptides are used for diagnosis, prognosis,
 XX prevention and treatment of multiple sclerosis (in its various forms and
 XX phases). They may also be useful in cases of e.g. Alzheimer's and
 XX Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 XX polyarthritis and lupus erythematosus, including use as vaccines and in
 XX gene therapy (expression of sense or antisense sequences). They can also
 XX be used to assess efficacy of potential therapeutic agents, particularly
 XX compounds that reduce or inhibit toxicity towards glial cells
 XX
 XX Sequence 189 AA;
 XX
 XX Query Match 95.2%; Score 79; DB 4; Length 189;
 XX Best Local Similarity 93.8%; Pred. No. 3.3e-05;
 XX Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YSLPKSEFAVPDLELP 16
 |||||
 Db 141 YSLPKSEFVVPDLELP 156
 |||||

```

RESULT 6
AAB31901
ID AAB31901 standard; protein; 193 AA.
XX
AC AAB31901;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roecklin D, Kolbe H, Charles M, Maicus C, Santoro L, Perron H;
XX
DR WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 162-163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 95.2%; Score 79; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 3.4e-05;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
Db ||||| |||||
145 YSLPKSEFAVPDLELP 160

RESULT 7
AAB31902
ID AAB31902 standard; protein; 193 AA.
XX
AC AAB31902;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roecklin D, Kolbe H, Charles M, Maicus C, Santoro L, Perron H;
XX
DR WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 162-163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 95.2%; Score 79; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 3.4e-05;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
Db ||||| |||||
145 YSLPKSEFAVPDLELP 160

```

```

DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200105422-A2.
XX
PD 25-JAN-2001.
XX
PF 17-JUL-2000; 2000WO-FR002057.
XX
PR 15-JUL-1999; 99FR-00009372.
XX
PA (INMR ) BIOMERIEUX STELHYS.
XX
PI Roecklin D, Kolbe H, Charles M, Maicus C, Santoro L, Perron H;
XX
DR WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 163; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 193 AA;
Query Match 95.2%; Score 79; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 3.4e-05;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
Db ||||| |||||
145 YSLPKSEFAVPDLELP 160

RESULT 8
AAB31904
ID AAB31904 standard; protein; 193 AA.
XX
AC AAB31904;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX

```

OS Homo sapiens.
 XX WO200105422-A2.
 PN PD 25-JAN-2001.
 XX PF 17-JUL-2000; 2000WO-FR002057.
 XX PR 15-JUL-1999; 99FR-00009372.
 XX PA (INMR) BIOMERIEUX STELHYS.
 XX PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX Claim 1; Page 164-165; 209pp; French.
 XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 XX Query Match 95.2%; Score 79; DB 4; Length 193;
 Best Local Similarity 93.8%; Pred. No. 3.4e-05;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 YSLPKSEFAVPDLELP 16
 Db 145 YSLPKSEFAVPDLELP 160
 RESULT 9
 AAB31896
 ID AAB31896 standard; protein; 193 AA.
 XX AC AAB31896;
 XX DT 15-MAY-2001 (first entry)
 XX DE Amino acid sequence of the human ganglioside GM2 activator protein.
 XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX OS Homo sapiens.
 XX WO200105422-A2.
 PN PD 25-JAN-2001.
 XX PF 17-JUL-2000; 2000WO-FR002057.

PR 15-JUL-1999; 99FR-00009372.
 XX (INMR) BIOMERIEUX STELHYS.
 XX PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 XX N-PSDB; AAF54698.
 XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX Claim 23; Page 158-159; 209pp; French.
 XX The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 XX Query Match 95.2%; Score 79; DB 4; Length 193;
 Best Local Similarity 93.8%; Pred. No. 3.4e-05;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 YSLPKSEFAVPDLELP 16
 Db 145 YSLPKSEFAVPDLELP 160
 RESULT 10
 AAB31928
 ID AAB31928 standard; protein; 193 AA.
 XX AC AAB31928;
 XX DT 15-MAY-2001 (first entry)
 XX DE Amino acid sequence of the human ganglioside GM2 activator protein.
 XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX OS Homo sapiens.
 XX WO200105422-A2.
 XX DT 25-JAN-2001.
 XX PF 17-JUL-2000; 2000WO-FR002057.
 XX PR 15-JUL-1999; 99FR-00009372.
 XX PA (INMR) BIOMERIEUX STELHYS.
 XX PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.

XX Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
XX PS Disclosure; Fig 1; 209pp; French.
XX
XX The present sequence represents a human polypeptide, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
XX SQ Sequence 193 AA;
Query Match 95.2%; Score 79; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 3.4e-05;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 YSLPKSEFAVPDLELP 16
||||| |||||
Db 145 YSLPKSEFVVPDLELP 160
RESULT 11
AAB31903
ID AAB31903 standard; protein; 193 AA.
AC AAB31903;
DT 15-MAY-2001 (first entry)
XX Amino acid sequence of a human protein.
XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
XX Homo sapiens.
XX WO200105422-A2.
XX 25-JAN-2001.
XX 17-JUL-2000; 2000WO-FR002057.
XX 15-JUL-1999; 99FR-00009372.
XX (INMR) BIOMERIEUX STELHYS.
XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
XX Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
XX Claim 1; Page 164; 209pp; French.
XX
XX The present sequence represents a human protein, which is used in the

CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
XX SQ Sequence 193 AA;
Query Match 95.2%; Score 79; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 3.4e-05;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 YSLPKSEFAVPDLELP 16
||||| |||||
Db 145 YSLPKSEFVVPDLELP 160
RESULT 12
ABG00720
ID ABG00720 standard; protein; 193 AA.
AC ABG00720;
XX 13-FEB-2002 (first entry)
XX Novel human diagnostic protein #711.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS64907.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 31079; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 193 AA;
 Query Match 95.2%; Score 79; DB 4; Length 193;
 Best Local Similarity 93.8%; Pred. No. 3.4e-05;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YSLPKSEFAVPDLELP 16
 |||||
 Db 145 YSLPKSEFAVPDLELP 160
 RESULT 13
 ABG31345
 ID ABG31345 standard; protein; 193 AA.
 AC ABG31345;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Human GM2 activator protein.
 XX
 KW Human; GM2 activator protein; ganglioside; platelet activating factor;
 KW PAF; inflammatory disorder; inflammatory bowel disease; asthma;
 KW autoimmune disease; lupus; hypersensitivity infection; rheumatism;
 KW rheumatoid arthritis; vasculitis; allergy; rhinitis; gout;
 KW tissue-specific condition; glomerulonephritis; hepatitis; redness;
 KW swelling; pain; polymorphonuclear leukocyte accumulation; virucide;
 KW antiinflammatory; antiasthmatic; antiarthritic; antirheumatic;
 KW antiallergic; hepatotropic; nephrotropic; immunosuppressive;
 KW tranquiliser.
 XX
 OS Homo sapiens.
 XX
 PN US6423680-B1.
 XX
 PD 23-JUL-2002.
 XX
 PF 30-OCT-1998; 98US-00183841.
 XX
 PR 30-OCT-1998; 98US-00183841.
 XX
 PA (HSCR-) HSC RES & DEV LP.
 XX
 PI Rigat B, Reynaud D, Mahuran D;
 XX
 DR WPI; 2002-664636/71.
 XX
 PT Composition useful for treating inflammatory conditions e.g. asthma
 PT comprises GM2 activator protein or GM2 activator peptide in combination
 PT with a carrier.
 XX
 PS Claim 3; Fig 1; 1lpp; English.
 XX
 CC The present invention relates to a composition comprising GM2
 CC (ganglioside) activator protein or a GM2 activator peptide derived from
 CC the GM2 activator protein in combination with a carrier. The composition
 CC comprises the protein or peptide in an amount of 1-100 mg. The GM2
 CC activator protein is capable of inhibiting platelet activating factor
 CC (PAF). The composition of the invention is useful for treating
 CC inflammatory disorders e.g. inflammatory bowel disease, asthma,

CC autoimmune disease (such as lupus), hypersensitivity infection,
 CC rheumatism (e.g. rheumatoid arthritis), vasculitis, allergies, rhinitis,
 CC gout and tissue-specific conditions (e.g. glomerulonephritis and
 CC hepatitis). The composition is capable of inhibiting platelet activating
 CC factor, is non-toxic, is efficacious and presents less severe side
 CC effects, including redness, swelling, pain and polymorphonuclear
 CC leukocyte accumulation at the inflammatory site and other associated
 CC cellular responses. The present sequence represents human GM2 activator
 CC protein
 XX
 SQ Sequence 193 AA;
 Query Match 95.2%; Score 79; DB 5; Length 193;
 Best Local Similarity 93.8%; Pred. No. 3.4e-05;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YSLPKSEFAVPDLELP 16
 |||||
 Db 145 YSLPKSEFAVPDLELP 160
 RESULT 14
 ABP65212
 ID ABP65212 standard; protein; 193 AA.
 AC ABP65212;
 XX
 DT 12-NOV-2002 (first entry)
 XX
 DE Hypoxia-regulated protein #86.
 XX
 KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;
 KW antiinflammatory; vulnery; gynecological; ophthalmological; vaccine;
 KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
 KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;
 KW preclapsia; atherosclerosis; inflammatory condition; wound healing;
 KW inflammation; erythropoiesis; hair loss; human.
 XX
 OS Homo sapiens.
 XX
 PN WO200246465-A2.
 XX
 PD 13-JUN-2002.
 XX
 PF 10-DEC-2001; 2001WO-GB005458.
 XX
 PR 08-DEC-2000; 2000GB-00030076.
 XX
 PR 08-FEB-2001; 2001GB-00003156.
 XX
 PR 25-OCT-2001; 2001GB-00025666.
 XX
 PA (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX
 PI White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;
 PI Rayner WN;
 XX
 DR WPI; 2002-627238/67.
 XX
 PT Identifying a gene involved in disease for treating hypoxia-regulated
 PT conditions, comprises comparing the transcriptome/proteome of two cell
 PT types under different conditions and identifying a differentially
 PT regulated gene.
 XX
 PS Claim 35; Page 397; 538pp; English.
 XX
 CC The present invention relates to methods for identifying genes and
 CC proteins that are implicated in a specific disease or physiological
 CC condition. The method comprises comparing the transcriptome/proteome of a
 CC specialised cell type implicated in a disease or condition with that of a
 CC second specialised cell type, under two experimental conditions, and
 CC identifying a gene that is differentially regulated in the two
 CC specialised cell types under experimental conditions. ABV77873-ABV78116
 CC and ABP65061-ABP65257 were identified using the methods of the invention.
 CC The coding sequences and proteins are useful for treating a disease in a

CC patient, for manufacture of a medicament for treating hypoxia-regulated
 CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
 CC biological response to hypoxia conditions, or hypoxic-associated
 CC pathology in a patient. The coding sequences and proteins are also useful
 CC for monitoring the therapeutic treatment of a disease or physiological
 CC condition, such as cancer, ischaemic conditions, reperfusion injury,
 CC retinopathy, neonatal stress, pre-eclampsia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss
 CC
 XX Sequence 193 AA;

Query Match 95.2%; Score 79; DB 5; Length 193;
 Best Local Similarity 93.8%; Pred. No. 3.4e-05;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
 |||||
 Db 145 YSLPKSEFVVPDLELP 160

RESULT 15

ADN95858
 ID ADN95858 standard; protein; 193 AA.

XX AC
 XX ADN95858;

XX 01-JUL-2004 (first entry)

XX Human BEC/LEC-related protein sequence SeqID782.

XX growth; differentiation; blood endothelial cell; BEC;
 KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
 KW lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;
 KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
 KW inflammatory disease; cancer metastasis; lymphatic system; human.

XX Homo sapiens.

XX WO2003080640-A1.

XX 02-OCT-2003.

XX 07-MAR-2003; 2003WO-US006900.

XX 07-MAR-2002; 2002US-0363019P.

XX (LUDW-) LUDWIG INST CANCER RES.
 XX (LICN) LICENTIA LTD.

XX Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;

XX WPI; 2003-876899/81.

XX N-PSDB; ADN95859.

PS Example 1; SEQ ID NO 782; 176pp; English.

XX This invention relates to a method of differentially modulating the
 CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
 CC endothelial cells (LEC) comprises contacting endothelial cells with a
 CC composition comprising an agent that differentially modulates blood or
 CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
 CC identifying a human subject with lymphoedema and with a mutation in at
 CC least one allele of a gene encoding a LEC protein, where the mutation
 CC correlates with lymphoedema in human subjects, and with the proviso that
 CC the LEC protein is not VEGFR-3; and administering to the subject a
 CC composition comprising a lymphatic growth agent selected from VEGF-C or
 CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
 CC the development of compounds with an antiangiogenic, cytostatic,
 CC vasotropic or antiinflammatory activity or for gene therapy. The method
 CC is useful in modulating the growth or differentiation of blood
 CC endothelial cells or lymphatic endothelial cells, in treating hereditary
 CC lymphoedema, in screening for an endothelial cell disorder or
 CC predisposition to the disorder or in monitoring the efficacy or toxicity

CC of a drug on endothelial cells. The agent is useful in manufacturing a
 CC medicament for the differential modulation of blood vessel endothelial
 CC cell or lymphatic vessel endothelial cell growth or differentiation. The
 CC lymphatic growth agent may also be used in manufacturing a medicament for
 CC the treatment of hereditary lymphoedema resulting from a mutation in a
 CC LEC gene or of other diseases involving the lymphatic vessels, such as
 CC various inflammatory diseases and cancer metastasis via the lymphatic
 CC system. The present sequence is that of a human LEC/BEC differentially
 CC expressed protein which is related to the method of the invention. Note:
 CC This sequence does not appear in the specification but was obtained by
 CC the indexer using the source data given in table 14 of the specification.
 XX
 SQ Sequence 193 AA;

Query Match 95.2%; Score 79; DB 7; Length 193;
 Best Local Similarity 93.8%; Pred. No. 3.4e-05;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
 |||||
 Db 145 YSLPKSEFVVPDLELP 160

Search completed: November 17, 2004, 07:10:03
 Job time : 6.65196 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 06:43:37 ; Search time 1.13357 Seconds
(without alignments)
936.062 Million cell updates/sec

Title: US-10-030-937-72
Perfect score: 83
Sequence: 1 YSLPKSEFAVPDLELP 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PGTUS COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	95.2	178	4	US-09-183-841-2
2	79	95.2	193	4	US-09-183-841-1
3	47	56.6	2710	2	US-08-568-459A-12
4	47	56.6	2710	2	US-08-487-826B-12
5	47	56.6	2710	3	US-09-210-288-12
6	47	56.6	3060	2	US-08-487-826B-14
7	44	53.0	307	4	US-09-543-681A-4241
8	43	51.8	471	3	US-08-911-853-4
9	43	51.8	471	3	US-09-479-409-4
10	43	51.8	471	3	US-09-479-453-4
11	43	51.8	812	4	US-09-489-039A-14282
12	41	49.4	186	4	US-09-107-532A-5213
13	41	49.4	226	4	US-09-107-532A-6759
14	41	49.4	366	4	US-09-252-931A-26920
15	41	49.4	510	3	US-08-948-564-4
16	40.5	48.8	476	4	US-09-248-796A-20470
17	40	48.2	112	4	US-09-489-039A-14284
18	40	48.2	251	4	US-09-248-796A-19708
19	40	48.2	4536	4	US-09-180-422B-27
20	40	48.2	4536	4	US-09-108-006C-1
21	40	48.2	4563	4	US-09-538-092-842
22	40	48.2	4563	4	US-09-252-991A-21629
23	39.5	47.6	548	4	US-09-079-030-76
24	39	47.0	40	4	US-09-489-039A-14205
25	39	47.0	580	4	US-09-252-991A-17570
26	39	47.0	776	4	US-09-107-532A-5284
27	39	47.0	900	4	US-09-107-532A-5284

Sequence 2, Appli
Sequence 79, Appl
Sequence 16766, A
Sequence 42023, A
Sequence 535, App
Sequence 19425, A
Sequence 1088, Ap
Sequence 17, Appl
Sequence 5905, Ap
Sequence 6932, Ap
Sequence 4, Appl
Sequence 214, App
Sequence 29833, A
Sequence 4853, Ap
Sequence 4, Appl
Sequence 5569, Ap
Sequence 31176, A
Sequence 6973, Ap

28 39 5 46 4 213 3 US-09-094-148-2
29 38 45 8 22 4 US-09-079-030-79
30 38 45 8 337 4 US-09-252-991A-16766
31 38 45 8 349 4 US-09-270-767-42023
32 38 45 8 421 4 US-09-198-452A-535
33 38 45 8 423 4 US-09-248-796A-19425
34 38 45 8 433 4 US-09-538-092-1088
35 38 45 8 537 3 US-08-886-886-17
36 38 45 8 547 4 US-09-107-532A-5905
37 38 45 8 564 4 US-09-543-681A-6932
38 38 45 8 697 2 US-08-674-351-4
39 38 45 8 840 4 US-09-079-030-214
40 37 5 45 2 170 4 US-09-252-991A-29833
41 37 5 45 2 454 3 US-09-134-001C-4853
42 37 5 45 2 1049 4 US-10-018-730A-4
43 37 44 6 252 4 US-09-328-352-6569
44 37 44 6 334 4 US-09-270-767-33176
45 37 44 6 405 4 US-09-328-352-6973

ALIGNMENTS

RESULT 1
US-09-183-841-2
; Sequence 2, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: His tag at residues 1 to 17
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: amino acid
; OTHER INFORMATION: sequence of GM2 protein using His6 tag
US-09-183-841-2

Query Match 95.2%; Score 79; DB 4; Length 178;
Best Local Similarity 93.8%; Pred. No. 1.6e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YSLPKSEFAVPDLELP 16
| | | | | | | | | | | | | | | | | | | | | |
Db 130 YSLPKSEFAVPDLELP 145

RESULT 2
US-09-183-841-1
; Sequence 1, Application US/09183841
; Patent No. 6423680
; GENERAL INFORMATION:
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: A No. 6423680el Inhibitor of Platelet Activating Factor
; FILE REFERENCE: vanz0010
; CURRENT APPLICATION NUMBER: US/09/183,841
; CURRENT FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL

```

; LOCATION: (33)..(55)
; FEATURE:
; OTHER INFORMATION: residues 56-63 are included in a further precursor
; OTHER INFORMATION: form of the protein
US-09-183-841-1

Query Match          95.2%; Score 79; DB 4; Length 193;
Best Local Similarity 93.8%; Pred. No. 1.7e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
Db 145 YSLPKSEFAVPDLELP 160

RESULT 3
US-08-568-459A-12
; Sequence 12, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobb Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-568-459A-12

Query Match          56.6%; Score 47; DB 2; Length 2710;
Best Local Similarity 64.3%; Pred. No. 26;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LPKSEFAVPDLELP 16
Db 2121 LPKNDGTVPDLEKP 2134

US-08-568-459A-12

RESULT 4
US-08-487-826B-12
; Sequence 12, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobb Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-487-826B-12

Query Match          56.6%; Score 47; DB 2; Length 2710;
Best Local Similarity 64.3%; Pred. No. 26;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LPKSEFAVPDLELP 16
Db 2121 LPKNDGTVPDLEKP 2134

US-09-210-288-12
; Sequence 12, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/210,288
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORGANISM: Plasmodium falciparum
 US-09-210-288-12

Query Match 56.6%; Score 47; DB 3; Length 2710;
 Best Local Similarity 64.3%; Pred. No. 26;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LPKSEFAVPDLELP 16
 ||||: |||||
 Db 2121 LPKNDGTVPDLEKP 2134

RESULT 6
 US-08-487-826B-14
 ; Sequence 14, Application US/08487826B
 ; Patent No. 5993827
 ; GENERAL INFORMATION:
 ; APPLICANT: Sim, Kim L.
 ; APPLICANT: Chitnis, Chetan
 ; APPLICANT: Miller, Louis H.
 ; APPLICANT: Peterson, David S.
 ; APPLICANT: Su, Xin-zhaun
 ; APPLICANT: Wellem, Thomas E.
 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe Martens Olson & Bear
 ; STREET: 620 Newport Center Drive 16th Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,826B
 ; FILING DATE: 10-SEP-1993
 ; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israelsen, Ned
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH121.001CPI
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3060 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-487-826B-14

Query Match 56.6%; Score 47; DB 2; Length 3060;
 Best Local Similarity 64.3%; Pred. No. 30;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LPKSEFAVPDLELP 16
 ||||: |||||
 Db 2119 LPKNDGTVPDLEKP 2132

RESULT 7
 US-09-543-681A-4241
 ; Sequence 4241, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 4241
 ; LENGTH: 307
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-4241

Query Match 53.0%; Score 44; DB 4; Length 307;
 Best Local Similarity 50.0%; Pred. No. 6.4;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLE 14
 ||||: |||||
 Db 220 YSLPQPTDPDIQ 233

RESULT 8
 US-08-911-853-4
 ; Sequence 4, Application US/08911853
 ; Patent No. 6048710
 ; GENERAL INFORMATION:
 ; APPLICANT: Gerritse, Gijbert
 ; APPLICANT: Quax, Wilhelmus J.
 ; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
 ; TITLE OF INVENTION: EXPRESSION LEVELS
 ; NUMBER OF SEQUENCES: 37
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genencor International
 ; STREET: 925 Page Mill Road
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304-1013
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA: US/08/911,853
APPLICATION NUMBER: US/08/911,853
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,092
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-911-853-4

Query Match 51.8%; Score 43; DB 3; Length 471;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 YSLPKSEFAVPDLELP 16
Db 41 YSIPTFDLVVSDLRP 56

RESULT 9

US-09-479-409-4
Sequence 4, Application US/09479409
Patent No. 6225106

GENERAL INFORMATION:
APPLICANT: Gerritse, Gijbert
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/479,409
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911,853

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
US-09-479-409-4

Query Match 51.8%; Score 43; DB 3; Length 471;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 YSLPKSEFAVPDLELP 16
Db 41 YSIPTFDLVVSDLRP 56

RESULT 10

US-09-479-453-4
Sequence 4, Application US/09479453
Patent No. 6313283

GENERAL INFORMATION:
APPLICANT: Gerritse, Gijbert
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/479,453
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911,853

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-479-453-4

Query Match 51.8%; Score 43; DB 3; Length 471;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 YSLPKSEFAVPDLELP 16
Db 41 YSIPTFDLVVSDLRP 56

RESULT 11

US-09-489-039A-14282
Sequence 14282, Application US/09489039A
Patent No. 6610836

GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001

```
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 14282
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14282

Query Match      51.8%   Score 43;   DB 4;   Length 812;
Best Local Similarity 66.7%   Pred. No. 32;
Matches 8;   Conservative 2;   Mismatches 2;   Indels 0;   Gaps 0;

QY      5 KSBFAVPDLELP 16
Db      762 KADFRVPPLELP 773

RESULT 12
US-09-107-532A-5213
; Sequence 5213, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5213:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...186
; SEQUENCE DESCRIPTION: SEQ ID NO: 5213:
US-09-107-532A-5213

Query Match      49.4%   Score 41;   DB 4;   Length 186;
Best Local Similarity 64.3%   Pred. No. 12;

; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 14282
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14282

Query Match      51.8%   Score 43;   DB 4;   Length 812;
Best Local Similarity 66.7%   Pred. No. 32;
Matches 8;   Conservative 2;   Mismatches 2;   Indels 0;   Gaps 0;

QY      5 KSBFAVPDLELP 16
Db      762 KADFRVPPLELP 773

RESULT 13
US-09-107-532A-6759
; Sequence 6759, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6759:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEtical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...226
; SEQUENCE DESCRIPTION: SEQ ID NO: 6759:
US-09-107-532A-6759

Query Match      49.4%   Score 41;   DB 4;   Length 226;
Best Local Similarity 50.0%   Pred. No. 15;
Matches 6;   Conservative 4;   Mismatches 2;   Indels 0;   Gaps 0;

QY      2 SLPKSFAVPDLEL 15
Db      37 SLPIRFAAPDCAL 50

RESULT 14
US-09-252-991A-26920
; Sequence 26920, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
```

1 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
2 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
3 FILE REFERENCE: 107196.136
4 CURRENT APPLICATION NUMBER: US/09/252,991A
5 CURRENT FILING DATE: 1999-02-18
6 PRIOR APPLICATION NUMBER: US 60/074,788
7 PRIOR FILING DATE: 1998-02-18
8 PRIOR APPLICATION NUMBER: US 60/094,190
9 PRIOR FILING DATE: 1998-07-27
10 NUMBER OF SEQ ID NOS: 33142
11 SEQ ID NO 26920
12 LENGTH: 366
13 TYPE: PRT
14 ORGANISM: Pseudomonas aeruginosa
15 US-09-252-991A-26920

Db 452 PGITFAIPNTELP 464

Search completed: November 17, 2004, 07:27:58
Job time : 2.22447 secs

Query Match 49.4%; Score 41; DB 4; Length 366;
Best Local Similarity 43.8%; Pred. No. 28;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
186 YTIPTFDMVSDRLP 201

RESULT 15
US-08-948-564-4
Sequence 4, Application US/08948564
Patent No. 6121512
GENERAL INFORMATION:
APPLICANT: Siminszky, Balazs
APPLICANT: Dewey, Ralph E.
APPLICANT: Corbin, Frederick T.
TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and
TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: PO Box 37428
CITY: Raleigh
STATE: No. 6121512th Carolina
COUNTRY: USA
ZIP: 27627
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,564
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 5051-409
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-854-1400
TELEFAX: 919-854-1401
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-948-564-4

Query Match 49.4%; Score 41; DB 3; Length 510;
Best Local Similarity 53.8%; Pred. No. 41;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 PKSEFAVPDLELP 16
| | | | | | | | | |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 07:01:42 ; Search time 3.82191 Seconds
(without alignments)
1482.515 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 83

Sequence: 1 YSLPKSEFAVPDLELP 16

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 1570615 segs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	95.2	193	14	US-10-170-385-389
2	47	56.6	2710	13	US-10-153-273-12
3	44.5	53.6	120	15	US-10-424-599-284187
4	44	53.0	328	14	US-10-369-493-4173
5	44	53.0	329	15	US-10-320-797-3110
6	43.5	52.4	1383	14	US-10-021-955-82
7	43.5	52.4	1383	14	US-10-021-955-86
8	43.5	52.4	1389	14	US-10-021-955-79
9	43.5	52.4	1389	14	US-10-021-955-87
10	43	51.8	183	17	US-10-425-115-261209
11	43	51.8	828	15	US-10-282-122A-60144
12	42.5	51.2	104	17	US-10-425-115-194142
13	42.5	51.2	1391	14	US-10-021-955-85

Sequence 1007, Ap
Sequence 77, Appl
Sequence 88, Appl
Sequence 10, Appl
Sequence 30, Appl
Sequence 29, Appl
Sequence 314, App
Sequence 232729,
Sequence 208031,
Sequence 234, App
Sequence 65806, A
Sequence 69294, A
Sequence 55243, A
Sequence 204585,
Sequence 140470,
Sequence 140467,
Sequence 39728, A
Sequence 235166,
Sequence 136391,
Sequence 356159,
Sequence 107307,
Sequence 321334,
Sequence 61811, A
Sequence 224622,
Sequence 208030,
Sequence 208027,
Sequence 19438, A
Sequence 8691, Ap
Sequence 30, Appl
Sequence 146878,
Sequence 53781, A
Sequence 248147,

ALIGNMENTS

RESULT 1

US-10-170-385-389
; Sequence 389, Application US/10170385
; Publication No. US20030203372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Jonathan
; APPLICANT: White, Katie Mary
; APPLICANT: Binley, William Nigel
; APPLICANT: Rayner, Stuart
; APPLICANT: Naylor, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 53268200100
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ IDS NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 389
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-170-385-389
Query Match 95.2%; Score 79; DB 14; Length 193;
Best Local Similarity 93.8%; Pred. No. 4.4e-05;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 YSLPKSEFAVPDLELP 16

Db 145 YSLPKSEFAVPDLELP 160
|||||

RESULT 2
US-10-153-273-12
; Sequence 12, Application US/10153273
; Publication No. US20020169305A1
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; Chitnis, Chetan
; Miller, Louis H.
; Peterson, David S.
; Su, Xin-zhaun
; Welles, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION/DOCKET NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-153-273-12

Query Match 56.6%; Score 47; DB 13; Length 2710;
Best Local Similarity 64.3%; Pred. No. 2e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LPKSEFAVPDLELP 16
|||: |||
Db 2121 LFKNDGTVPDLKP 2134

RESULT 3
US-10-424-599-284187
; Sequence 284187, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 284187
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_98646C.1.pep
US-10-424-599-284187

Query Match 53.6%; Score 44.5; DB 15; Length 120;
Best Local Similarity 68.8%; Pred. No. 15;
Matches 11; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
QY 1 YSLPKSEFAVPDLELP 16
|||: |||
Db 98 YFLPLSE-SEPDLELP 112

RESULT 4
US-10-369-493-4173
; Sequence 4173, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 4173
LENGTH: 328
TYPE: PRT
ORGANISM: Neurospora crassa
FEATURE:
NAME/KEY: unsure
LOCATION: (1) (328)
OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-4173

Query Match 53.0%; Score 44; DB 14; Length 328;
Best Local Similarity 61.5%; Pred. No. 58;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 PKSEFAVPDLELP 16
|||: |||
Db 105 PKAEWKVPTLSLP 117

RESULT 5
US-10-320-797-3110
; Sequence 3110, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroshkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797


```

; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3110
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Cryptococcus neoformans
US-10-320-797-3110

Query Match      53.0%; Score 44; DB 15; Length 329;
Best Local Similarity 56.2%; Pred. No. 59;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
Db 58 FSLPVKEFIIDUFLP 73

RESULT 6
US-10-021-955-82
; Sequence 82, Application US/10021955
; Publication No. US20030039987A1
; GENERAL INFORMATION:
; APPLICANT: Lupski, James R
; APPLICANT: Boerkoel, Cornelius F
; APPLICANT: Takashima, Hiroshi
; TITLE OF INVENTION: Defects in Periaxin Associated with Myelinopathies
; FILE REFERENCE: P02086USI/10026309
; CURRENT APPLICATION NUMBER: US/10/021,955
; PRIOR FILING DATE: 2001-12-13
; PRIOR APPLICATION NUMBER: US 60/255,217
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 82
; LENGTH: 1383
; TYPE: PRT
; ORGANISM: Rat
US-10-021-955-82

Query Match      52.4%; Score 43.5; DB 14; Length 1383;
Best Local Similarity 60.0%; Pred. No. 3.7e+02;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 3 LPKS-EFAVPDLELP 16
Db 453 LPKAPAAIPDVQLP 467

RESULT 7
US-10-021-955-86
; Sequence 86, Application US/10021955
; Publication No. US20030039987A1
; GENERAL INFORMATION:
; APPLICANT: Lupski, James R
; APPLICANT: Boerkoel, Cornelius F
; APPLICANT: Takashima, Hiroshi
; TITLE OF INVENTION: Defects in Periaxin Associated with Myelinopathies
; FILE REFERENCE: P02086USI/10026309
; CURRENT APPLICATION NUMBER: US/10/021,955
; PRIOR FILING DATE: 2001-12-13
; PRIOR APPLICATION NUMBER: US 60/255,217
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 86
; LENGTH: 1383
; TYPE: PRT
; ORGANISM: Rat
US-10-021-955-86

Query Match      52.4%; Score 43.5; DB 14; Length 1383;
Best Local Similarity 60.0%; Pred. No. 3.7e+02;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 3 LPKS-EFAVPDLELP 16
Db 453 LPKAPAAIPDVQLP 467

RESULT 8
US-10-021-955-87
; Sequence 87, Application US/10021955
; Publication No. US20030039987A1
; GENERAL INFORMATION:
; APPLICANT: Lupski, James R
; APPLICANT: Boerkoel, Cornelius F
; APPLICANT: Takashima, Hiroshi
; TITLE OF INVENTION: Defects in Periaxin Associated with Myelinopathies
; FILE REFERENCE: P02086USI/10026309
; CURRENT APPLICATION NUMBER: US/10/021,955
; PRIOR FILING DATE: 2001-12-13
; PRIOR APPLICATION NUMBER: US 60/255,217
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 87
; LENGTH: 1383
; TYPE: PRT
; ORGANISM: Rat
US-10-021-955-87

Query Match      52.4%; Score 43.5; DB 14; Length 1383;
Best Local Similarity 60.0%; Pred. No. 3.7e+02;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 3 LPKS-EFAVPDLELP 16
Db 453 LPKAPAAIPDVQLP 467

RESULT 9
US-10-021-955-79
; Sequence 79, Application US/10021955
; Publication No. US20030039987A1
; GENERAL INFORMATION:
; APPLICANT: Lupski, James R
; APPLICANT: Boerkoel, Cornelius F
; APPLICANT: Takashima, Hiroshi
; TITLE OF INVENTION: Defects in Periaxin Associated with Myelinopathies
; FILE REFERENCE: P02086USI/10026309
; CURRENT APPLICATION NUMBER: US/10/021,955
; PRIOR FILING DATE: 2001-12-13
; PRIOR APPLICATION NUMBER: US 60/255,217
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79
; LENGTH: 1389
; TYPE: PRT
; ORGANISM: Rat
US-10-021-955-79

Query Match      52.4%; Score 43.5; DB 14; Length 1389;
Best Local Similarity 60.0%; Pred. No. 3.7e+02;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 3 LPKS-EFAVPDLELP 16
Db 453 LPKAPAAIPDVQLP 467

RESULT 10
US-10-425-115-261209

```

```
; Sequence 261209, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 261209
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_169837C.1.pep
US-10-425-115-261209

Query Match      51.8%; Score 43; DB 17; Length 183;
Best Local Similarity 72.7%; Pred. No. 44;
Matches      8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      2 SLPKSEFAVPD 12
      |||:|||||
Db      25 SLPRSFVPD 35

RESULT 11
US-10-282-122A-60144
; Sequence 60144, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60144
; LENGTH: 828
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-60144

Query Match      51.8%; Score 43; DB 15; Length 828;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches      8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      5 KSEFAVPDLELP 16
      ||:|||||
Db      758 KADFRVPPLELP 769

RESULT 12
US-10-425-115-194142
; Sequence 194142, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 194142
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_108637C.1.pep
US-10-425-115-194142

Query Match      51.2%; Score 42.5; DB 17; Length 104;
Best Local Similarity 38.1%; Pred. No. 28;
Matches      8; Conservative 5; Mismatches 3; Indels 5; Gaps 1;

QY      1 YSLPKSEFA-----VPDLELP 16
      ::|||:|||||
Db      64 FEIPSEYSHNTIIVDPDEIP 84

RESULT 13
US-10-021-955-85
; Sequence 85, Application US/10021955
; Publication No. US20030039987A1
; GENERAL INFORMATION:
; APPLICANT: Lupski, James R
; APPLICANT: Boerkoel, Cornelius F
; APPLICANT: Takashima, Hiroshi
; TITLE OF INVENTION: Defects in Periaxin Associated with Myelinopathies
; FILE REFERENCE: P02086USI/10026309
; CURRENT APPLICATION NUMBER: US/10/021,955
; CURRENT FILING DATE: 2001-12-13
; PRIOR APPLICATION NUMBER: US 60/255,217
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 85
; LENGTH: 1391
; TYPE: PRT
; ORGANISM: Mouse
US-10-021-955-85

Query Match      51.2%; Score 42.5; DB 14; Length 1391;
Best Local Similarity 66.7%; Pred. No. 5.4e+02;
Matches     10; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
```

Search completed: November 17, 2004, 07:38:24
Job time : 5.82191 secs

QY 3 LPK-SEFAVPDLELP 16
||| |||||:
Db 510 LPKVPMAVPDVHLP 524

RESULT 14

US-10-408-765A-1007
; Sequence 1007, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Wainock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1007
; LENGTH: 1398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1007

Query Match 51.2%; Score 42.5; DB 16; Length 1398;
Best Local Similarity 66.7%; Pred. No. 5.5e+02;
Matches 10; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 3 LPK-SEFAVPDLELP 16
||| |||||:
Db 607 LPKVPMAVPDVHLP 621

RESULT 15

US-09-940-227-77
; Sequence 77, Application US/09940227
; Publication No. US20030017468A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Sei Yu
; APPLICANT: Macina, Roberto
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Hervé
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific
; FILE REFERENCE: DEX-0230
; CURRENT APPLICATION NUMBER: US/09/940,227
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 60/228,378
; PRIOR FILING DATE: 2000-08-28
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 77
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-940-227-77

Query Match 51.2%; Score 42.5; DB 10; Length 1461;
Best Local Similarity 66.7%; Pred. No. 5.7e+02;
Matches 10; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 3 LPK-SEFAVPDLELP 16
||| |||||:
Db 605 LPKVPMAVPDVHLP 619

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 06:36:49 ; Search time 0.81146 Seconds
(without alignments)
1897.160 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 83

Sequence: 1 YSLPKSEFAVPDLELP 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	95.2	162	2 S13195	ganglioside M2 act
2	79	95.2	193	2 I54178	ganglioside M2 act
3	79	95.2	200	2 S22411	ganglioside M2 act
4	68	81.9	193	2 S35613	ganglioside M2 act
5	47	56.6	3078	2 T28432	variant-specific s
6	44	53.0	353	2 S62854	MG068 homolog D02
7	44	53.0	493	2 B48362	transcription init
8	44	53.0	683	2 D82574	TonB-dependent rec
9	43.5	52.4	1389	2 I58157	peritaxin - rat
10	43	51.8	157	2 A72662	hypothetical prote
11	42	50.6	199	2 E81930	hypothetical prote
12	42	50.6	216	1 F69586	orotate phosphorib
13	42	50.6	400	2 D71520	probable transamin
14	42	50.6	467	2 S53950	probable membrane
15	41	49.4	478	2 C83055	probable two-compo
16	41	49.4	490	2 S00995	gene twist protein
17	41	49.4	510	2 T05939	cytochrome P450 mo
18	40.5	48.8	231	1 C69798	probable membrane
19	40.5	48.8	262	2 A98355	transcription regu
20	40	48.2	186	2 T00395	hypothetical prote
21	40	48.2	253	2 T41418	40S ribosomal prot
22	40	48.2	264	2 P00478	pistil extensin-li
23	40	48.2	275	2 E60950	apolipoprotein B-1
24	40	48.2	393	2 P00479	pistil extensin-li
25	40	48.2	401	2 T51407	proline-rich prote
26	40	48.2	504	2 S51590	mitochondrial proc
27	40	48.2	559	2 AC1512	hypothetical prote
28	40	48.2	723	2 T32136	hypothetical prote
29	40	48.2	861	2 B04085	probable transcrip

30 40 48.2 886 2 T39229
31 40 48.2 962 2 S03818
32 40 48.2 4563 1 LPHUB
33 39.5 47.6 453 2 B83242
34 39.5 47.6 1134 2 F87642
35 39 47.0 65 2 S35172
36 39 47.0 158 2 H87643
37 39 47.0 166 2 E96607
38 39 47.0 168 2 T27563
39 39 47.0 275 2 T42747
40 39 47.0 279 2 A95398
41 39 47.0 288 2 D87222
42 39 47.0 311 2 D95265
43 39 47.0 332 2 E90293
44 39 47.0 774 2 D83208
45 39 47.0 899 2 B86812

ALIGNMENTS

RESULT 1

S13195

ganglioside M2 activator protein - human

C;Species: Homo sapiens (man)

C;Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004

C;Accession: S13195

R;Fuerst, W.; Schubert, J.; Machleidt, W.; Meyer, H.E.; Sandhoff, K.

Eur. J. Biochem. 192, 709-714, 1990

A;Title: The complete amino-acid sequences of human ganglioside GM2 activator protein and

A;Reference number: S13195; MUID:91006185; PMID:2209618

A;Accession: S13195

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-162 <PUE>

A;Cross-references: UNIPROT:P17900

Query Match 95.2%; Score 79; DB 2; Length 162;

Best Local Similarity 93.8%; Pred. No. 3.6e-06;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16

Db 114 YSLPKSEFVVPDLELP 129

RESULT 2

I54178

ganglioside M2 activator protein precursor - human

C;Species: Homo sapiens (man)

C;Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004

C;Accession: I54178; JQ1037; S05036; S22410; S17107

R;Xie, B.; Kennedy, J.L.; McInnes, B.; Auger, D.; Mahuran, D.

Genomics 14, 796-798, 1992

A;Title: Identification of a processed pseudogene related to the functional gene encoding

A;Reference number: I54178; MUID:93052421; PMID:1427911

A;Accession: I54178

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-193 <RES>

A;Cross-references: UNIPROT:P17900; GB:I01439; NID:g183358; PIDN:AAA52767.1; PID:g183359

R;Xie, B.; McInnes, B.; Neote, K.; Lamhonwah, A.M.; Mahuran, D.

Biochem. Biophys. Res. Commun. 177, 1217-1223, 1991

A;Title: Isolation and expression of a full-length cDNA encoding the human GM2 activator

A;Reference number: JQ1037; MUID:91282768; PMID:2059210

A;Accession: JQ1037

A;Molecule type: mRNA

A;Residues: 1-18, 'A', 20-193 <XIE>

A;Cross-references: GB:M76477; NID:g183356; PIDN:AAA35907.1; PID:g183357

A;Experimental source: Hela cell

A;Note: 19-Thr and 69-Met were also found

R;Schroeder, M.; Klima, H.; Nakano, T.; Kwon, H.; Quintern, L.E.; Gaertner, S.; Suzuki, T.

FEBS Lett. 251, 197-200, 1989
A;Title: Isolation of a cDNA encoding the human G(M2) activator protein.
A;Reference number: S05036; MUID:89325664; PMID:2753159
A;Accession: S05036
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 15-193 <SCH>
A;Cross-references: GDB:X16087; NID:g31852; PIDN:CAA34215.1; PID:g31853
R:Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A;Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A;Reference number: S22410; MUID:92207171; PMID:1554364
A;Accession: S22410
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 43-142, 'VST', <NAG>
A;Cross-references: EMBL:X61094; NID:g31854; PIDN:CAA43407.1; PID:g31855
C;Comment: This protein transports GM2 ganglioside from the lysosomal membrane to hexosa
C;Genetics:
A;Gene: GDB:GM2A
A;Cross-references: GDB:120000; OMIM:272750
A;Map position: 5q32-5q33
F:1-23/Domain: signal sequence #status predicted <SIG>
F:32-193/Product: GM2 ganglioside activator protein #status predicted <GM2>
F:32-193/Product: ganglioside M2 activator #status predicted <MAT>

Query Match 95.2%; Score 79; DB 2; Length 193;
Best Local Similarity 93.8%; Pred. No. 4.4e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
DB 145 YSLPKSEFAVPDLELP 160
|||||

RESULT 3
S22411
ganglioside M2 activator protein (clone pGAP2) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
C;Accession: S22411; S17108
R:Nagarajan, S.; Chen, H.C.; Li, S.C.; Li, Y.T.; Lockyer, J.M.
Biochem. J. 282, 807-813, 1992
A;Title: Evidence for two cDNA clones encoding human GM2-activator protein.
A;Reference number: S22410; MUID:92207171; PMID:1554364
A;Accession: S22411
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-200 <NAG>
A;Cross-references: UNIPROT:P17900; EMBL:X61095; NID:g31856; PIDN:CAA43408.1; PID:g31857

Query Match 95.2%; Score 79; DB 2; Length 200;
Best Local Similarity 93.8%; Pred. No. 4.6e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
DB 152 YSLPKSEFAVPDLELP 167
|||||

RESULT 4
S35613
ganglioside M2 activator protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 07-May-1999
C;Accession: S35613
R:Bellachioma, G.; Stirling, J.L.; Oriacchio, A.; Beccari, T.
Biochem. J. 294, 227-230, 1993
A;Title: Cloning and sequence analysis of a cDNA clone coding for the mouse G(M2) activa
A;Reference number: S35613; MUID:93371367; PMID:7689629
A;Accession: S35613
A;Status: preliminary
A;Molecule type: mRNA

Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLLEL 15
||| :||| :|||
Db 96 YKIPADFAVLELEL 110

RESULT 7
B48362
transcription initiation factor sigma 54 - Alcaligenes eutrophus
N;Alternate names: rpoN protein
C;Species: Alcaligenes eutrophus
C;Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: B48362; S24624
R;Warrelmann, J.; Eitinger, M.; Schwartz, E.; Romermann, D.; Friedrich, B.
Arch. Microbiol. 158, 107-114, 1992
A;Title: Nucleotide sequence of the rpoN (hno) gene region of Alcaligenes eutrophus: evl
A;Reference number: A48362; MUID:93038001; PMID:1417413
A;Accession: B48362
A;Molecule type: DNA
A;Residues: 1-493 <WAR>
A;Cross-references: UNIPROT:P28615; EMBL:X66793; NID:G38783; PIDN:CAA47287.1; PID:G38785
A;Note: sequence extracted from NCBI backbone (NCBIP:115740)
C;Genetics:
A;Gene: rpoN; hno
C;Superfamily: Pseudomonas transcription initiation factor sigma
C;Keywords: DNA binding; nucleotidyltransferase; sigma factor; transcription initiation

Query Match 53.0%; Score 44; DB 2; Length 493;
Best Local Similarity 53.8%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDL 13
||| :||| :|||
Db 281 YSRPEADFVVDV 293

RESULT 8
D82674
TonB-dependent receptor for iron transport XF1496 [imported] - Xylella fastidiosa (strain
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: D82674
R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: D82674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-683 <SIM>
A;Cross-references: UNIPROT:Q9PD83; GB:AEO03979; GB:AEO03849; NID:G9106520; PIDN:AAF8430
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Mencia, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1496

Query Match 53.0%; Score 44; DB 2; Length 683;
Best Local Similarity 50.0%; Pred. No. 20;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLLEL 16
||| :||| :|||
Db 558 YSRQKAEIAPDPSVP 573

RESULT 9
I58157
peritaxin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 15-Mar-2004
C;Accession: I58157
R;Gillespie, C.S.; Sherman, D.L.; Blair, G.B.; Brophy, P.J.
Neuron 12, 497-508, 1994
A;Title: Peritaxin, a novel protein of myelinating Schwann cells with a possible role in
A;Reference number: I58157; MUID:94206531; PMID:8155317
A;Accession: I58157
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1389 <RES>
A;Cross-references: EMBL:Z29649; NID:G505296; PIDN:CAA82757.1; PID:G505297
F;25-94/Domain: GLGF domain homology <GLG3>

Query Match 52.4%; Score 43.5; DB 2; Length 1389;
Best Local Similarity 60.0%; Pred. No. 55;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 3 LPKS-EFAVPDLLEL 16
||| :||| :|||
Db 453 LPKAPAAIPDVQLP 467

RESULT 10
A72662
hypothetical protein APE0721 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: A72662
R;Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takaha
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Ki
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: A72662
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-157 <KAW>
A;Cross-references: UNIPROT:Q9YE49; DBJ:AP000060; NID:G5104188; PIDN:BAA79697.1; PID:G51
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0721
C;Superfamily: Aeropyrum pernix hypothetical protein APE0721

Query Match 51.8%; Score 43; DB 2; Length 157;
Best Local Similarity 53.3%; Pred. No. 5.6;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLLEL 15
||| :||| :|||
Db 63 FGIPVSIFKLPDLLEL 77

RESULT 11
E81930
hypothetical protein NMA0849 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C;Accession: E81930
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A:Reference number: A81775; MUID:202222556; PMID:10761919
 A:Accession: E81930
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-199 <PAR>
 A:Cross-references: UNIPROT:Q9JVG5; GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB8413
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: NMA0849
 C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0226

Query Match 50.6%; Score 42; DB 2; Length 199;
 Best Local Similarity 50.0%; Pred. No. 11;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 LPKSEFAVPDLELP 16
 |||||:|
 Db 33 LPQSEFSIQCPEP 46

RESULT 12

F69686
 orotate phosphoribosyltransferase (EC 2.4.2.10) pyrE [similarity] - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 02-Jul-1998 #text_change 09-Jul-2004
 A:Accession: F69686; A30492
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C:Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
 A:; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: F69686
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-216 <KUN>

A:Cross-references: UNIPROT:P25972; GB:Z99112; GB:AL009126; NID:G2633902; PIDN:CAB13430.
 A:Experimental source: strain 168
 R:Quinn, C.L.; Stephenson, B.T.; Switzer, R.L.
 J. Biol. Chem. 266, 9113-9127, 1991

A:Title: Functional organization and nucleotide sequence of the Bacillus subtilis pyrim
 A:Reference number: A39845; MUID:91225016; PMID:1709162

A:Accession: A30492

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 'M', 8-216 <QUI>

A:Cross-references: GB:M59757; NID:g387576

A:Note: these authors used TTG as a start codon

C:Genetics:

A:Gene: pyrE

C:Function:

A:Description: catalyzes the formation of orotidine 5'-phosphate and pyrophosphate from
 A:Pathway: pyrimidine nucleotide biosynthesis
 C:Superfamily: orotate phosphoribosyltransferase; orotate phosphoribosyltransferase hom
 C:Keywords: glycosyltransferase; pentosyltransferase; pyrimidine nucleotide biosynthesis
 F:7-200/Domain: orotate phosphoribosyltransferase homology <OPT>

Query Match 50.6%; Score 42; DB 1; Length 216;
 Best Local Similarity 56.2%; Pred. No. 12;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
 |||||:|

Db 158 YGLPKAEFAFAKAEPL 173

RESULT 13

D71520
 probable transaminase (EC 2.6.1.-) aspC [similarity] - Chlamydia trachomatis (serotype D)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
 A:Accession: D71520
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
 A:Reference number: A71570; MUID:9900809; PMID:9784136
 A:Accession: D71520
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-400 <ARN>
 A:Cross-references: UNIPROT:O84395; GB:AE001312; GB:AE001273; NID:g3328812; PIDN:AAC6798
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:

A:Gene: aspC

C:Superfamily: aspartate transaminase

C:Keywords: aminotransferase

Query Match 50.6%; Score 42; DB 2; Length 400;

Best Local Similarity 72.7%; Pred. No. 24;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 SLPKSEFAVPD 12

|||||:|

Db 219 SLPKSIFRIPD 229

RESULT 14

S53950
 probable membrane protein YMR033w - Yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein YMR973.07
 C:Species: Saccharomyces cerevisiae
 C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
 A:Accession: S53950
 R:Bowman, S.
 submitted to the EMBL Data Library, May 1995

A:Reference number: S53944

A:Accession: S53950

A:Molecule type: DNA

A:Residues: 1-467 <BOW>

A:Cross-references: UNIPROT:Q05123; EMBL:Z49213; NID:g798952; PID:g798959; MIPS:YMR033w

A:Experimental source: strain AB972

C:Genetics:

A:Gene: SGD.ARP9

A:Cross-references: SGD:S0004636; MIPS:YMR033w

A:Map position: 13R

A:Introns: 10/3

C:Keywords: transmembrane protein

F:76-92/Domain: transmembrane #status predicted <TM>

Query Match 50.6%; Score 42; DB 2; Length 467;

Best Local Similarity 43.8%; Pred. No. 29;

Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16

:|:|:|:|

Db 23 FGLNEETFTVPELEIP 38

|||||:|

RESULT 15

C83055

probable two-component response regulator PA4726 [imported] - Pseudomonas aeruginosa (st
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 02-Jun-2003
 A:Accession: C83055
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,


```

.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: C83055
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-478 <STO>
A;Cross-references: GB:AE004886; GB:AE004091; NID:g9950982; PIDN:AAG08112.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA4726
C;Superfamily: response regulator, NtrC type; response regulator homology; RNA polymeras

Query Match          49.4%; Score 41; DB 2; Length 478;
Best Local Similarity 43.8%; Pred. No. 44;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      1 YSLPKSEFAVPDLRLP 16
      ||::|: |||||
Db      41 YTIPSFDMVVSRLRLP 56

```

Search completed: November 17, 2004, 07:24:56
Job time : 2.81146 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 06:33:32 ; Search time 4.63957 Seconds
(without alignments)
1984.236 Million cell updates/sec

Title: US-10-030-937-72

Perfect score: 83

Sequence: 1 YSLPKSEFAVPDLELP 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listings first 45 summaries

Database : Uniprot 02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	95.2	189	2 Q6LBL5	Q6Lb15 homo sapien
2	79	95.2	189	2 CAA43994	Caa43994 homo sapi
3	79	95.2	193	1 SAP3 HUMAN	PL17900 homo sapien
4	72	86.7	190	2 Q8HXX6	Q8hxx6 macaca fasc
5	69	83.1	199	2 Q6IN37	Q6in37 rattus norv
6	69	83.1	199	2 Q8CJH4	Q8cjb4 rattus norv
7	68	81.9	193	1 SAP3 MOUSE	Q60648 mus musculu
8	60	72.3	146	2 Q6Q7X4	Q6q7x4 felis silve
9	60	72.3	146	2 AAS64351	Aas64351 felis sil
10	60	72.3	151	2 Q6Q7X5	Q6q7x5 felis silve
11	60	72.3	151	2 AAS64350	Aas64350 felis sil
12	49	59.0	894	2 Q8GSR4	Q8gsf4 oryza sativ
13	47	56.6	200	2 Q83FA3	Q83fa3 coxiella bu
14	47	56.6	1801	2 Q8WSJ2	Q8wsj2 bombyx mori
15	47	56.6	3078	2 Q26031	Q26031 plasmodium
16	45	54.2	265	2 Q7YQW3	Q7yqw3 notoryctes
17	45	54.2	265	2 Q7YQW5	Q7yqw5 dendrotylus
18	45	54.2	266	2 Q72D26	Q72d26 desulfovibr
19	45	54.2	266	2 AAS95585	Aas95585 desulfovi
20	44	53.0	197	2 Q6GLN6	Q6gln6 xenopus lae
21	44	53.0	353	1 YF91 MYCPN	Y50336 mycoplasma
22	44	53.0	492	2 Q9EX69	Q9ex69 alcaligenes
23	44	53.0	493	1 RP54 ALCEU	R28615 alcaligenes
24	44	53.0	683	2 Q9P0B3	Q9pd83 xylella fas
25	44	53.0	786	2 Q9NEP7	Q9nef7 drosophila
26	43.5	52.4	1383	1 PRAX RAT	Q63425 rattus norv
27	43	51.8	149	2 Q6WMT5	Q6wmt5 branchiosto
28	43	51.8	149	2 AAQ83886	Aaq83886 branchios
29	43	51.8	157	2 Q9YE49	Q9ye49 aeropyrum p
30	43	51.8	231	2 Q7YQV7	Q7yqv7 dasynurus al
31	43	51.8	253	2 Q7YQV9	Q7yqv9 sminthopsis

32	43	51.8	265	2 Q7YQW0	Q7yqw0 phascogale
33	43	51.8	942	2 Q8CX25	Q8cx25 streptococc
34	43	51.8	942	2 Q8DXI6	Q8dxyl6 streptococc
35	42.5	51.2	1391	1 PRAX MOUSE	O5103 mus musculu
36	42.5	51.2	1391	2 Q6NVF7	Q6nvf7 mus musculu
37	42.5	51.2	1391	2 AAH68135	Aah68135 mus muscu
38	42.5	51.2	1461	1 PRAX HUMAN	Q9bxm0 homo sapien
39	42.5	51.2	1461	2 AAH67286	Aah67286 homo sapi
40	42	50.6	72	2 Q997F7	Q997f7 tt virus. o
41	42	50.6	121	2 Q85624	Q85624 mycobacteri
42	42	50.6	199	1 HAM1 NEIMA	Q9jvg5 neisseria m
43	42	50.6	216	1 PYRE BACSU	P25972 bacillus su
44	42	50.6	254	2 Q976Z3	Q976z3 sulfolobus
45	42	50.6	255	2 Q7YQW8	Q7yqw8 caluromys p

ALIGNMENTS

RESULT 1					
Q6Lb15	Q6Lb15	PRELIMINARY;	PRT;	189 AA.	
ID	Q6Lb15				
AC	Q6Lb15				
DT	05-JUL-2004 (Tremblrel. 27, Created)				
DT	05-JUL-2004 (Tremblrel. 27, Last sequence update)				
DT	03-JUL-2004 (Tremblrel. 27, Last annotation update)				
DB	GM2 activator protein.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92008637; PubMed=1915857;				
RA	Klima H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,				
RA	Sandhoff K.				
RT	"Characterization of full-length cDNAs and the gene coding for the				
RT	human GM2 activator protein."				
RL	FEBS Lett. 289:260-264 (1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93277527; PubMed=8503891;				
RA	Klima H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,				
RA	Sandhoff K.				
RT	"Over-expression of a functionally active human GM2-activator protein				
RT	in escherichia coli."				
RL	Biochem. J. 292:571-576 (1993).				
DR	EMBL; X62078; CAA43994.1; -				
DR	InterPro; IPR003172; EI_DerP2_DerF2.				
DR	SMART; SM00737; ML; 1.				
SQ	SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;				
Query Match 95.2%; Score 79; DB 2; Length 189;					
Best Local Similarity 93.8%; Pred. No. 4.3e-05;					
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
Qy	1 YSLPKSEFAVPDLELP 16				
Db	141 YSLPKSEFVVPDLELP 156				
RESULT 2					
CAA43994	CAA43994	PRELIMINARY;	PRT;	189 AA.	
ID	CAA43994				
AC	CAA43994;				
DT	02-MAR-2004 (Tremblrel. 27, Created)				
DT	02-MAR-2004 (Tremblrel. 27, Last sequence update)				
DT	02-MAR-2004 (Tremblrel. 27, Last annotation update)				
DB	GM2 activator protein.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				

RN SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klinka H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
human GM2 activator protein.";
RL FEBS Lett. 289:260-264(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277527; PubMed=8503891;
RA Klinka H., Klein A., Van Echten G., Schwarzmann G., Suzuki K.,
Sandhoff K.;
RT "Over-expression of a functionally active human GM2-activator protein
in *Escherichia coli*.";
RL Biochem. J. 292:571-576(1993).
DR EMBL; X62078; CAA43994.1; -;
SQ SEQUENCE 189 AA; 20362 MW; 9B8C7F18DC7439BE CRC64;
Query Match 95.2%; Score 79; DB 2; Length 189;
Best Local Similarity 93.8%; Pred. No. 4.3e-05;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 YSLPKSEFAVPDLELP 16
||||| |||||
Db 141 YSLPKSEFVVPDLELP 156
RESULT 3
SAP3_HUMAN STANDARD; PRT; 193 AA.
AC P17500; Q14426; Q14428;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate
activator protein) (Shingolipid activator protein 3) (SAP-3).
GN Name=GM2A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RX MEDLINE=91282768; PubMed=2059210;
RA Xie B., McInnes B., Neote K., Lamhonwah A.-M., Mahuran D.;
RT "Isolation and expression of a full-length cDNA encoding the human G-
M2 activator protein.";
RL Biochem. Biophys. Res. Commun. 177:1217-1223(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92008637; PubMed=1915857;
RA Klinka H., Tanaka A., Schnabel D., Nakano T., Schroeder M., Suzuki K.,
Sandhoff K.;
RT "Characterization of full-length cDNAs and the gene coding for the
human GM2 activator protein.";
RL FEBS Lett. 289:260-264(1991).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANT ALA-19.
RX MEDLINE=92207171; PubMed=1554364;
RA Nagarajan S., Chen H.C., Li S.C., Li Y.T., Lockyer J.;
RT "Evidence for two cDNAs encoding human GM2-activator protein.";
RL Biochem. J. 282:807-813(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93052421; PubMed=1427911;
RA Xie B., Kennedy J.L., McInnes B., Auger D., Mahuran D.J.;
RT "Identification of a processed pseudogene related to the functional
gene encoding the GM2 activator protein: localization of the
pseudogene to human chromosome 3 and the functional gene to human
chromosome 5.";
RL Genomics 14:796-798(1992).

RN SEQUENCE FROM N.A.
RX MEDLINE=99294584; PubMed=10364519;
RA Chen B., Rigat B., Curry C., Mahuran D.J.;
RT "Structure of the GM2A gene: identification of an exon 2 nonsense
mutation and a naturally occurring transcript with an in-frame
deletion of exon 2.";
RL Am. J. Hum. Genet. 65:77-87(1999).
RN [6]
RP SEQUENCE FROM N.A.
RX TISSUE=Uterus;
RA MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klauser R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [7]
RP SEQUENCE OF 15-193 FROM N.A.
RX MEDLINE=89325664; PubMed=2753159;
RA Schroeder M., Klinka H., Nakano T., Kwon H., Quintern L.E.,
Gaertner S., Suzuki K., Sandhoff K.;
RT "Isolation of a cDNA encoding the human GM2 activator protein.";
RL FEBS Lett. 251:197-200(1989).
RN [8]
RP SEQUENCE OF 32-193.
RX TISSUE=Kidney;
RA MEDLINE=91006165; PubMed=2209618;
RA Furst W., Schubert J., Machleidt W., Meyer H.E., Sandhoff K.;
RT "The complete amino-acid sequences of human ganglioside GM2 activator
protein and cerebroside sulfate activator protein.";
RL Eur. J. Biochem. 192:709-714(1990).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.0 ÅNGSTROMS).
RX MEDLINE=20545600; PubMed=11090283; DOI=10.1006/jmbi.2000.4225;
RA Wright C.S., Li S.-C., Rastinejad F.;
RT "Crystal structure of human GM2-activator protein with a novel beta-
cup topology.";
RL J. Mol. Biol. 304:411-422(2000).
RN [10]
RP VARIANT TSD-AB ARG-138.
RX MEDLINE=92008638; PubMed=1915858;
RA Schroeder M., Schnabel D., Suzuki K., Sandhoff K.;
RT "A mutation in the gene of a glycolipid-binding protein (GM2
activator) that causes GM2-gangliosidosis variant AB.";
RL FEBS Lett. 290:1-3(1991).
RN [11]
RP VARIANT TSD-AB PRO-169.
RX MEDLINE=94063850; PubMed=8244332;
RA Schroeder M., Schnabel D., Hurwitz R., Young E., Suzuki K.,
Sandhoff K.;
RT "Molecular Genetics of GM2-gangliosidosis AB variant: a novel mutation
and expression in BHK cells.";
RL Hum. Genet. 92:437-440(1993).
RN [12]
RP VARIANT TSD-AB LYS-88 DEL.
RX MEDLINE=97055887; PubMed=8900233;
RA Schepers U., Glombitza G., Lemm T., Hoffmann A., Chabas A., Ozand P.,

RA Sandhoff K.;
 RT "Molecular analysis of a GM2-activator deficiency in two patients with
 RL GM2-gangliosidosis AB variant.,"
 RL Am. J. Hum. Genet. 59:1048-1056(1996).
 CC -1- FUNCTION: Binds gangliosides and stimulates ganglioside GM2
 CC degradation. It stimulates only the breakdown of ganglioside GM2
 CC and glycolipid GA2 by beta-hexosaminidase A. It extracts single
 CC GM2 molecules from membranes and presents them in soluble form to
 CC beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and
 CC conversion to GM3.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- DISEASE: Defects in GM2A are the cause of Tay-Sachs disease AB
 CC variant (TSD-AB) [MIM:272750]; also known as GM2-gangliosidosis
 CC type AB.
 CC -1- DATABASE: NAME=GM2AB; NOTE=GM2A mutation database;
 CC WWW="http://www.hexadb.mcgill.ca/?Topic=GM2AB&Page=MutationSubmission".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 DR ENBL; M76477; AAA35907.1; --
 DR ENBL; X62078; CAA43993.1; --
 DR ENBL; X61095; CAA43408.1; ALT INIT.
 DR ENBL; L01439; AAA52767.1; --
 DR ENBL; AF124719; AAD25741.1; --
 DR ENBL; AF124717; AAD25741.1; JOINED.
 DR ENBL; AF124718; AAD25741.1; JOINED.
 DR ENBL; BC009273; AAH09273.1; --
 DR ENBL; X16087; CAA34215.1; --
 DR PIR; I54178; I54178.
 DR PIR; S13195; S13195.
 DR PIR; S22411; S22411.
 DR PDB; 1G13; X-ray; A/B/C=32-193.
 DR Genew; HGNC:4367; GM2A.
 DR MIM; 272750; --
 DR GO; GO:0005764; C:lysosome; NAS.
 DR GO; GO:0010290; F:sphingolipid activator protein activity; NAS.
 DR GO; GO:0019377; P:glycolipid catabolism; NAS.
 DR GO; GO:0030149; P:sphingolipid catabolism; NAS.
 DR InterPro; IPR003172; EL_DerP2_DerF2.
 DR SMART; SM00737; ML; 1.
 KW 3D-structure; Direct protein sequencing; Disease mutation;
 KW Glycoprotein; GM2-gangliosidosis; Lysosome; Polymorphism; Signal;
 KW Sphingolipid metabolism.
 FT SIGNAL 1 31
 FT CHAIN 32 193 Ganglioside GM2 activator.
 FT DISULFID 39 183
 FT DISULFID 99 106
 FT DISULFID 112 138
 FT DISULFID 125 136
 FT CARBOHYD 63 63
 FT VARIANT 19 19 N-linked (GlcNAc...).
 FT T -> A.
 FT /FTID=VAR_013830.
 FT Missing (in 80% of the protein).
 FT /FTID=VAR_006946.
 FT Missing (in TSD-AB).
 FT /FTID=VAR_011697.
 FT C -> R (in TSD-AB).
 FT /FTID=VAR_006947.
 FT R -> P (in TSD-AB).
 FT /FTID=VAR_011698.
 FT V -> I (in Ref. 3).
 FT V -> M (in Ref. 3).
 FT VARIANT 32 33
 FT VARIANT 88 88
 FT VARIANT 138 138
 FT VARIANT 169 169
 FT CONFLICT 59 59
 FT CONFLICT 69 69
 FT STRAND 35 38
 FT TURN 41 43
 FT STRAND 46 54
 FT STRAND 58 59
 FT STRAND 63 72

FT STRAND 76 76
 FT TURN 81 90
 FT STRAND 91 92
 FT STRAND 93 96
 FT STRAND 100 100
 FT TURN 101 102
 FT STRAND 103 103
 FT STRAND 107 108
 FT TURN 109 110
 FT HELIX 111 118
 FT TURN 121 122
 Query Match 95.2%; Score 79; DB 1; Length 193;
 Best Local Similarity 93.8%; Pred. No. 4.4e-05;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YSLPKSEFAVPDLELP 16
 DB 145 YSLPKSEFAVPDLELP 160
 RESULT 4
 Q8HXX6 PRELIMINARY; PRT; 190 AA.
 ID Q8HXX6
 AC Q8HXX6; 2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Ganglioside GM2 activator.
 GN Names=gm2a;
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OC NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain cerebellum cortex;
 RA Kusuda J., Osada N., Sugano S., Hashimoto K.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB083313; BAC20592.1; --
 DR HSP; P17900; IG13.
 DR InterPro; IPR003172; EL_DerP2_DerF2.
 DR SMART; SM00737; ML; 1.
 SQ SEQUENCE 190 AA; 20494 MW; 9F9582BBE75715C3 CRC64;
 Query Match 86.7%; Score 72; DB 2; Length 190;
 Best Local Similarity 87.5%; Pred. No. 0.00066;
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 YSLPKSEFAVPDLELP 16
 DB 142 YSLPKSEFAVPDLELP 157
 RESULT 5
 Q6IN37 PRELIMINARY; PRT; 199 AA.
 ID Q6IN37
 AC Q6IN37; 2004 (Tremblrel. 27, Created)
 DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
 DE GM2 ganglioside activator protein.
 GN Name=Gm2a;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 MEDLINE=22388257; PubMed=12477932;

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton M., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC072474; AAH72474.1; --
 DR InterPro: IPR003172; E1_DerP2_DerF2.
 DR SMART: SM00737; ML; 1.
 SQ SEQUENCE 199 AA; 21493 MW; C2B5203FDFABF507 CRC64;

Query Match 83.1%; Score 69; DB 2; Length 199;
 Best Local Similarity 81.2%; Pred. No. 0.0022; 3; Indels 0; Gaps 0;
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 YSLPKSEFAVPDLELP 16
 Db 151 YSLPGSNFTVPDLELP 166

RESULT 6
 Q8CJH4 PRELIMINARY; PRT; 199 AA.
 AC Q8CJH4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-NOV-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE GM2 activator protein.
 GN Name=rGM2AP;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Miwa N., Okada T., Nakamura S.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB051391; BAC24018.1; --
 DR HSP: P17900; IG33.
 DR InterPro: IPR003172; E1_DerP2_DerF2.
 DR SMART: SM00737; ML; 1.
 SQ SEQUENCE 199 AA; 21521 MW; D585203FDFABF507 CRC64;

Query Match 83.1%; Score 69; DB 2; Length 199;
 Best Local Similarity 81.2%; Pred. No. 0.0022; 3; Indels 0; Gaps 0;
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 YSLPKSEFAVPDLELP 16
 Db 151 YSLPGSNFTVPDLELP 166

RESULT 7
 SAP3_MOUSE STANDARD; PRT; 193 AA.
 ID SAP3_MOUSE

AC Q0648; Q61610; Q61819;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Ganglioside GM2 activator precursor (GM2-AP) (Cerebroside sulfate
 DE activator protein) (Shingolipid activator protein 3) (SAP-3).
 GN Name=Gm2a;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
 RX MEDLINE=95229165; PubMed=7713516;
 RA Yamanaka S., Johnson O.N., Iyu M.S., Kozak C.A., Proia R.L.;
 RT "The mouse gene encoding the GM2 activator protein (Gm2a): cDNA
 RT sequence, expression, and chromosome mapping.";
 RL Genomics 24:601-604(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA;
 RX MEDLINE=93371367; PubMed=7689829;
 RA Bellachio G., Stirling J.L., Oriacchio A., Beccari T.;
 RT "Cloning and sequence analysis of a cDNA clone coding for the mouse
 RT GM2 activator protein.";
 RL Biochem. J. 294:227-230(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA;
 RX MEDLINE=97224573; PubMed=9060405;
 RA Bertoni C., Appolloni M.G., Stirling J.L., Li S.C., Li Y.T.,
 RT "Structural organization and expression of the gene for the mouse GM2
 RT activator protein.";
 RL Mamm. Genome 8:90-93(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton M., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Binds gangliosides and stimulates ganglioside GM2
 CC degradation. It stimulates only the breakdown of ganglioside GM2
 CC and glycolipid GM2 by beta-hexosaminidase A. It extracts single
 CC GM2 molecules from membranes and presents them in soluble form to
 CC beta-hexosaminidase A for cleavage of N-acetyl-D-galactosamine and
 CC conversion to GM3.
 CC -!- SUBCELLULAR LOCATION: Lysosomal.
 CC -!- TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney and
 CC testis.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

AAS64350
 ID AAS64350 PRELIMINARY; PRT; 151 AA.
 AC AAS64350;
 DT 25-MAR-2004 (TREMBlrel. 27, Created)
 DT 25-MAR-2004 (TREMBlrel. 27, Last sequence update)
 DT 25-MAR-2004 (TREMBlrel. 27, Last annotation update)
 DE GM2 activator protein (Fragment).
 GN GM2A.
 OS *Felis silvestris catus* (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 OX NCBI_TaxID=9685;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Martin D.R., Cox N.R., Morrison N.E., Kenamer D.M., Peck S.L.,
 RA Dodson A.N., Gentry A.S., Griffin B., Rolima M.D., Baker H.J.,
 RT "Mutation of the GM2 activator protein in a feline model of GM2
 RT gangliosidosis";
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY553653; AAS64350.1; --
 FT NON TER
 SQ SEQUENCE 151 AA; 16310 MW; E23EF16CE5844A5F CRC64;

Query Match 72.3%; Score 60; DB 2; Length 151;
 Best Local Similarity 62.5%; Pred. No. 0.056;
 Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLELP 16
 |||||:|:|:|:|:|:|:
 Db 103 YSLPESDFTLPOLEVP 118

RESULT 12
 Q8GSP4 PRELIMINARY; PRT; 894 AA.
 AC Q8GSP4;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE SNF2P.
 OS Name=Snf2P;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]_TaxID=39947;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22359412; PubMed=12471446;
 RA Yan L., Echenique V., Busso C., SanMiguel P., Ramakrishna W.,
 RA Bennetzen J.L., Harrington S., Dubcovsky J.;
 RT "Cereal Genes similar to Snf2 define a new subfamily that includes
 RT human and mouse genes";
 RL Mol. Genet. Genomics 268:488-499 (2002).
 DR EMBL; AF459086; AANI4536.1; --
 DR EMBL; AF459087; AANI4537.1; --
 DR Gramene; Q8GSP4; --
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR009022; EFG III V.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR000330; SNF2_N.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00176; SNF2_N; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELICCC; 1.
 SQ SEQUENCE 894 AA; 106656 MW; 29A7CD823AD7BAAB CRC64;

Query Match 59.0%; Score 49; DB 2; Length 894;
 Best Local Similarity 61.5%; Pred. No. 29;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 LPKSEFAVPDLEL 15
 :|:|:|:|:|:|:
 Db 817 VPRSEISPDLEL 829

RESULT 13
 Q83FA3 PRELIMINARY; PRT; 200 AA.
 AC Q83FA3;
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)

DE HAM1 protein.
 GN OrderedLocusNames=CBU00043;
 OS Coccidia burnetii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
 OC Cocciliaceae; Coccidia.
 OX NCBI_TaxID=777;
 RN [1]_TaxID=777;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nine Mile phase I / RSA 493;
 RX MEDLINE=22608657; PubMed=12704232;
 RA Seshadri R., Paulsen I.T., Eissen J.A., Read T.D., Nelson K.E.,
 RA Nelson W.C., Ward N.L., Tettelin H., Daviden T.M., Beanan M.J.,
 RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
 RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
 RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.P.;
 RT "Complete genome sequence of the Q-fever pathogen, Coccidia
 RT burnetii";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460 (2003).
 CC -!- SIMILARITY: Belongs to the HAM1 NTPase family.
 DR EMBL; AB016960; AAO89612.1; --
 DR HSSP; Q57679; 2MJP.
 DR TIGR; CBU0043; --
 DR InterPro; IPR002637; Ham1p like.
 DR Pfam; PF01725; Ham1p like; 1.
 DR TIGRFAMs; TIGR00042; Ham1p like; 1.
 KW Complete proteome; Hydrolase.
 SQ SEQUENCE 200 AA; 21777 MW; A3SDF16EE7025113 CRC64;
 Query Match 56.6%; Score 47; DB 2; Length 200;
 Best Local Similarity 58.3%; Pred. No. 12;
 Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 3 LPKSEFAVPDLE 14
 :|:|:|:|:|:|:
 Db 29 IPQTEFSVPDIE 40

RESULT 14
 Q8WSJ2 PRELIMINARY; PRT; 1801 AA.
 AC Q8WSJ2;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Ovarian serine protease.
 GN Name=OSP;
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Bombycidae;
 OX NCBI_TaxID=7091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kendirgi F., Iatrou K.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 DR EMBL; AF234884; AAL62027.1; --
 DR HSSP; P00760; IEZX.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002172; LDL_receptor A.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase S1A.
 DR InterPro; IPR009003; Pept Ser Cys.
 DR InterPro; IPR001190; Sscr_receptor.
 DR Pfam; PF00057; Ldl_recept_a; 5.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.

Fri Nov 19 14:12:41 2004

```

DR SMART; SM00192; LDLa; 7.
DR SMART; SM00020; Tryp.SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 3.
DR PROSITE; PS00688; LDLRA_2; 6.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 2.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 1801 AA; 203641 MW; E5227772A342A411 CRC64;

Query Match      56.6%; Score 47; DB 2; Length 1801;
Best Local Similarity 56.2%; Pred.No. 1.4e+02;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 YSLPKSEFAVPDLLELP 16
DB 270 YELPHSEYAPQYESP 285

RESULT 15
Q26031 PRELIMINARY; PRT; 3078 AA.
AC Q26031;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Variant-specific surface protein.
DE Name=var-1;
GN Plasmodium falciparum.
OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DQ2;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Wellem T.E.;
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RT infected erythrocytes.";
RL Cell 82:89-100(1995).
RL EMBL; L40608; AAA75396.1; -.
DR PIR; T28432; T28432.
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:0005539; F:glycosaminoglycan binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR004258; PFEMP.
DR Pfam; PF05658; Hep Hag; 1.
DR Pfam; PF03011; PFEMP; 2.
DR SEQUENCE 3078 AA; 349299 MW; C8037C2BC3CCD7C3 CRC64;

Query Match      56.6%; Score 47; DB 2; Length 3078;
Best Local Similarity 64.3%; Pred.No. 2.5e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LPKSEFAVPDLLELP 16
DB 2119 LPKNDGTVPDLLEXP 2132

```

Search completed: November 17, 2004, 07:22:37
 Job time : 6.63957 secs

This Page Blank (uspio)